| DECLARATION OF CAMERON JENNINGS, PH.D. UNDER 37 C.F.R. §1.132 | Attorney Docket No. | 29729/38914 |
|---|---|---|
| | Confirmation No. | 2874 |
| | First Named Inventor | DUNN, ROSANNE D |
| | Application Number | 10/590,690 |
| | Filing Date | May 21, 2007 |
| Address to: Commissioner for Patents P.O. Box 1450 Alexandria, VA 22313-1450 | Group Art Unit | 1644 |
| | Examiner Name | Schwadron, R. |
| | Title: "TARGET FOR B-CELL DISORDERS " | |

Sir:

1. I, Cameron Jennings, PhD, declare and say I am a resident of 1 Janice Street, Warners Bay, New South Wales 2282, Australia.

2. I am an employee of Graham A Brown and Associates. My previous scientific employment consists of Immune System Therapeutics Ltd (Sydney), Harvard University (Boston) and La Trobe University (Melbourne). Details of my career as well as publications may be found in my curriculum vitae (Exhibit A).

3. I understand that US Patent Application No. 10/590,690 (referred to hereafter as the "Patent Application") is assigned to Immune System Therapeutics Limited. I presently own options which would allow me to purchase shares in Immune System Therapeutics Limited should the company become publicly listed.

4. I have been asked by FB Rice & Co, Patent Attorneys for Immune System Therapeutics Limited, to provide an independent opinion on the state of knowledge surrounding kappa and lambda light chains and on the invention described in the Patent Application. I have been asked in particular to comment on the obviousness rejections set out in the Office Action dated 11 September 2009 issued in connection with the Patent Application.

5. I have read the Patent Application and have reviewed the claims that I understand are presently being considered by the United States Patent and Trademark Office. I understand that the claimed invention (referred to hereafter as the "Invention") relates to methods of treating B-cell disorders by administering an antibody or ligand that specifically binds to lambda myeloma antigen (LMA). The importance of this invention is that it provides, for the first time, LMA as a target that is selective for tumor B-cells

expressing LMA. Antibodies and ligands that bind LMA do not target normal B-cells which express lambda light chain in the context of intact immunoglobulin.

6. I have considerable experience in the technical field of the claimed invention. I have over 3 years of experience in research involving immunoglobulin light chains and membrane-bound proteins. Prior to this time, my research consisted of a senior post-doctoral fellowship at Harvard University researching membrane bound malaria vaccine candidates and a PhD that was co-supervised by Professor Marilyn Anderson (Biochemistry Department at La Trobe University) and Professor David Craik (Institute of Molecular Bioscience at The University of Queensland).

7. I have reviewed the Office Action dated 11 September 2009 in relation to the above-referenced Patent Application. I understand that the Patent Office has taken the position that the Invention would be obvious to a person skilled in the art in light of Uhr *et al.* (US 7,792,447), Raison *et al.* (WO 03/004056) and Abe *et al.* (1993).

8. Uhr *et al.* are cited as disclosing antibodies that bind intact immunoglobulin associated lambda light chain on tumor cells. Raison *et al.* are cited as disclosing that malignant cells express both kappa and lambda light chains, that free kappa light chain is expressed on the cell surface of kappa light chain expressing myeloma cells and that antibodies which bind free kappa light chain can be used to treat tumors. Abe *et al.* are cited as disclosing antibodies that bind free light chain and not intact immunoglobulin.

9. It is my understanding that the Patent Application and the cited publications are to be viewed from the perspective of one of ordinary skill in the art in the relevant field (a "Skilled Person") at the time of filing of the Patent Application in question. I have been asked to consider this time to be the period around or before 27 February 2004 ("the Relevant Period"). I would expect a Skilled Person in the field of antibody therapy during the Relevant Period to have been represented by a scientist with a PhD degree in Biochemistry and/or at least 3 to 5 years experience in the field of Biochemistry, or an educational background at the same degree level in a related field and equivalent level of experience.

10. I am very familiar with the technical field of the claimed invention. I am qualified to analyze literature in this field and to provide my opinion as to what literature in this field discloses or suggests to the Skilled Person at the Relevant Period.

11. By the Relevant Period I had attained at least the level of such a Skilled Person, and further in view of my qualifications discussed above, I believe that I am qualified by training and experience to address what a Skilled Person would have understood from reading the Patent Application and the cited publications.

12. The Examiner's objection appears to be based on the premise that kappa and lambda light chains are the two known alleles of immunoglobulin light chain and therefore once free kappa light chain had been found expressed on the surface of myeloma cells, it would have been expected that free lambda light chains would also be expressed on the surface of malignant B cells.

13. In my view, it is not correct to assume that the expression and localisation of the kappa light chain in myeloma cells is in any way predictive of the expression and localisation of lambda light chains in malignant B cells.

14. Immunoglobulins are composed of two chains, light and heavy, which form a functional heterodimer. The light chain molecules are of two protein families, "kappa" and "lambda". The kappa and Lambda gene loci are located on chromosomes 2 and 22 respectively and differ in their number of variable (V), joining (J) and constant (C) genes as well as their general arrangement. For example, there is a single C kappa gene (Zimmer *et al.*, 1990) while there are at least 4 functional C lambda genes (Dariavach *et al.*, 1987). While both kappa and lambda light chains maintain a conserved role when present in intact immunoglobulin, no normal biological function has been attributed to light chains alone.

15. Although kappa and lambda light chains both comprise a conserved (C) and variable (V) domain (within the respective families) and they both complete a multimeric immunoglobulin, the proteins share minimal sequence identity within their C domains (Kabat *et al.*, 1975). Furthermore, the variable domains of the kappa and lambda light chains are derived from genetic recombination events and the recombined genes also undergo the process of somatic hypermutaion to give rise to antigen binding moieties (in the context of the whole immunoglobulin). As such the V regions vary in their sequence as well.

16. Recent information indicates that although kappa and lambda V domains have a typical Immunoglobulin fold consisting of two antiparallel β-barrel sheets, they diverge in the

processes they use to force the phenomenon known as the 'strand switch' (James *et al.*, 2007). This phenomenon is dependent on local structure and confirms the variability in sequence, structure and biological function between the two immunoglobulin light chains.

17. The immunoglobulin beta sandwich, referred to above, contains seven beta strands which form a sandwich of two beta sheets. This structural motif is found on a vast array of protein sub-families with diverse biological activities and sub-cellular locations (for a complete set of proteins in the SCOP data base, see (Murzin *et al.*, 1995; available from http://scop.mrc-lmb.cam.ac.uk/scop/data/scop.b.c.b.html). Proteins that fall within this structural family include Myelin oligodendrocyte glycoprotein (MOG), the T-cell antigen receptor and TREM-1 (triggering receptor expressed on myeloid cells) to name a few.

18. As further evidence of differences in structure of kappa and lambda light chain, the *Peptostreptococcus magnus* protein L binds kappa light chains but does not bind lambda light chains (Graille et al., 2001). This again suggests variation in sequence and local structure of the kappa and lambda light chains

19. The differences in the structures of the kappa and lambda light chains is further highlighted by observations indicating that lambda antibodies are found in two-thirds of light chain Amyloidosis cases, whereas kappa light chains mediate greater than 85% of Light-Chain Deposition Disease (LCDD). Thus, the light chain composition appears to confer different pathologies to kappa or lambda light chains (James *et al.*, 2007).

20. The difference in the structure of amyloid fibrils (fibrilar) and the LCDD deposits (amorphous) also reflects the difference in the general structure of kappa and lambda light chains (Khurana et al., 2001). Moreover, lambda light chains exist predominantly as dimers while kappa light chains are mostly present as monomers (Bradwell, 2008; Solomon and Weiss, 1995). As a consequence, the character and rate of the catabolic processes that are involved in the clearance of kappa and lambda light chains are different. It has been suggested that this phenomenon contributes to the predominance of lambda light chains in amyloidosis (Solomon and Weiss, 1995).

21. There exists three broad categories of surface expressed proteins: i) integrally associated proteins possessing hydrophobic surfaces that readily interact with the acyl core of the bilayer; ii) membrane proteins that are covalently attached to certain phospholipids; and iii) peripheral proteins that associate with the membrane through charge-charge

electrostatic interactions (Sachs and Engelman 2006). Examination of the sequence of kappa and lambda light chains does not reveal candidate residues suitable for covalent attachment to the membrane. Thus, it can be surmised that the membrane interaction of either kappa or lambda light chains would occur via hydrophobic and/or electrostatic interactions.

22. Differences in the primary sequence of kappa and lambda light chains affects not only the charge of the exposed side chains of the proteins but also their predominant presence as either monomers or dimers respectively. Thus, in view of the differences in primary sequence of kappa and lambda light chains, which in turn affects the presence/absence of exposed hydrophobic surfaces, and the lack of a membrane targeting sequence in these proteins, the expression of either of these two proteins on the cell surface could not be predicted. Thus, it could not have been predicted by analysis of the structure or sequence of kappa or lambda light chains that either of these proteins would associate with the membrane of malignant B cells.

23. In my opinion, therefore, there is nothing in the cited prior art to suggest that a Skilled Person could have predicted that free lambda light chain would be associated with the membrane of tumor B-cells. There was no suggestion nor motivation, therefore, around February 2004 for a Skilled Person to investigate free lambda light chains as a potential therapeutic target on tumor B-cells.

24. I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under § 1001 of Title XVIII of the United States Code, and that such willful false statements may jeopardize the validity of the Patent Application or any patent issuing thereon.

7th March 2010

Date

Name: Cameron Jennings, PhD

Chapter 13

# The kidney and monoclonal free light chains

*AR Bradwell, P Cockwell and C Hutchison*



**Summary:** Monoclonal serum free light chains:-

1. Cause renal impairment in approximatcly 30% of patients with MM and dialysis dependent renal failure in 10%.
2. Should be measured in all MM patients to identify those at risk of renal damage.
3. Are not adequately removed by plasma exchange.
4. Can be removed by haemodialysis using *"high cut-off"* dialysers, leading to renal recovery.

## 13.1. Introduction

Renal failure is a major cause of morbidity and mortality in patients with MM. At initial presentation, up to 50% of patients have renal impairment (serum creatinine >1.5mg/dL or >130μmol/L); 12 to 20% have acute renal failure (ARF) and 10% become dialysis dependent. They represent 2% of the dialysis population and there are approximately 5,000 new patients, worldwide, each year. Furthermore, there is a 50% mortality within 6 months of diagnosis *(Kyle 7, Hutchison 2)*

While reversible factors such as dehydration, hypercalcaemia and medication are frequently involved, monoclonal FLCs are the most potent cause of irreversible renal failure. Large amounts of sFLCs readily pass through the glomerular fenestrations and overwhelm the absorptive capacity of the proximal tubules. On entering the distal tubules, they co-precipitate with Tamm-Horsfall protein to form waxy casts that both block the flow of urine and cause interstitial inflammation *(Herrera 1)*. Furthermore, high concentrations of FLCs are directly toxic to tubular cells *(Sanders 1,2)*.

Studies have analyzed renal recovery rates after FLC removal by plasma exchange. This is a logical approach, but results have been disappointing. Although an early report was optimistic *(Zucchelli)*, the largest and most recent controlled trial (97 patients) showed no clinical benefit *(Clark WF)*. A subsequent editorial in the Journal of the American Society of Nephrology (JASN) listed the shortcomings of this study, including the failure to monitor either serum or urine FLC concentrations *(Ritz)*. It was noted,

*"This resembles anti-hypertensive treatment without measuring blood pressure."* Clearly, the efficiency of plasma exchange for FLC removal could not be judged.

Because FLCs are relatively small protein molecules (κ~25 kDa: dimeric λ~50 kDa) they are present in similar concentrations in serum, extravascular compartments and tissue oedema fluid *(Takagi)*. Thus, the intravascular compartment may contain only 15 to 20% of the total amount. A series of 3.5 litre plasma exchanges that removed 65% of intravascular FLCs on each occasion would have little overall impact, particularly if production were not reduced at the same time by chemotherapy. An alternative approach is to remove FLCs by haemodialysis. Although this is not possible with routine dialyzers because of their small pore sizes (12-15kDa), a new generation of *"high cut-off"* dialyzers, allows FLC removal *(Hutchison 2)*. By using extended dialysis, large amounts of FLCs can be removed without the attendant clotting and deproteination problems that may limit the extended use of plasma exchange.

This chapter discusses the normal renal handling of FLCs, their role in renal failure in MM, clinical case studies and current management strategies such as plasma exchange. A mathematical model of FLC removal is presented and plasma exchange is compared with the utility of haemodialysis. Finally, clinical evidence for the beneficial use of *"FLC removal haemodialysis"* is presented.

## 13.2. Normal free light chain clearance and metabolism

In normal individuals, sFLCs are rapidly cleared by the kidneys depending upon their molecular size *(Figure 13.1 and Chapter 3)*. Monomeric FLCs, characteristically κ, are cleared in 2-4 hours at 40% of the glomerular filtration rate. Dimeric FLCs, typically λ, are cleared in 3-6 hours at 20% of the glomerular filtration rate, while larger polymers are cleared more slowly. Removal is prolonged to 2-3 days in MM patients who are in complete renal failure when FLCs are removed by the liver and other tissues. In contrast, IgG has a normal serum half-life of 21 days that is not affected by renal

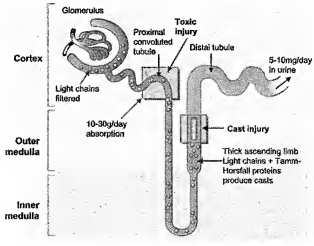



**Figure 13.1.** Renal injury caused by FLCs. (Courtesy of R Johnson and J Feehally).

impairment.

After filtration by the glomeruli, FLCs enter the proximal tubules and bind to brush-border membranes via low-affinity, high-capacity receptors called cubulins and megalins *(Johnson RJ)*. Binding provokes internalisation of the FLCs, subsequent proteolysis into smaller peptides and finally their excretion into the urine flow. The concentration of FLCs leaving the proximal tubules, therefore, depends upon the amount in the glomerular filtrate, competition for binding uptake from other proteins and the absorptive capacity of the tubular cells. A reduction in GFR, due to loss of nephrons, increases sFLC concentrations so that more is filtered by the remaining functioning nephrons. Subsequently, and with increasing renal failure, hyperfiltering glomeruli leak albumin and other proteins which compete with FLCs for absorption thereby causing more to enter the distal tubules.

FLCs entering the distal tubule can bind to uromucoid (Tamm-Horsfall protein). This is the predominant protein in normal urine and is thought to be important in preventing ascending urinary infections. It is a glycoprotein (85kDa) that aggregates into high molecular weight polymers of 20-30 units. Interestingly, it contains a short peptide motif that has a high affinity for FLCs *(Ying)*.

### 13.3. Nephrotoxicity of monoclonal free light chains

The main pathology in myeloma kidney is cast nephropathy. This is caused by precipitation of FLCs with uromucoid as waxy casts and is characteristically found in acute renal failure associated with MM *(Figures 13.2 and 13.3) (Johnson RJ)*. The casts obstruct tubular fluid flow, leading to disruption of the basement membrane and interstitial damage. Rising concentrations of sFLCs are filtered by the remaining functioning nephrons which become blocked, leading to a vicious cycle of further increases in sFLC concentrations and progressive renal damage. This may explain why some MM patients, without apparent pre-existing renal impairment, suddenly develop

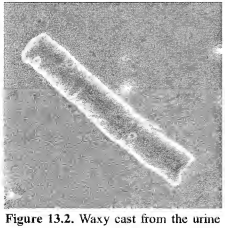


**Figure 13.2.** Waxy cast from the urine of a patient with multiple myeloma. (Courtesy of R Johnson and J Feehally).

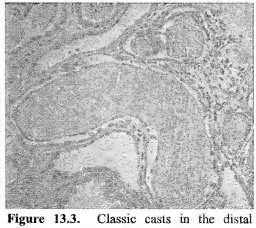


**Figure 13.3.** Classic casts in the distal tubules of a patient with light chain multiple myeloma. (Courtesy of C Hutchison).

*Proc. Natl. Acad. Sci. USA*
Vol. 84, pp. 9074–9078, December 1987
Genetics

# Human immunoglobulin $C_\lambda 6$ gene encodes the Kern$^+$Oz$^-$ λ chain and $C_\lambda 4$ and $C_\lambda 5$ are pseudogenes

(isotypes/allotypes/Bence Jones proteins/constant region genes)

P. DARIAVACH, G. LEFRANC, AND M.-P. LEFRANC

Laboratoire d'Immunogénétique, Centre National de la Recherche Scientifique Unité Associée 1191, Génétique Moléculaire, Université des Sciences et Techniques du Languedoc, Place E. Bataillon, 34060 Montpellier Cedex, France

*Communicated by C. Milstein, August 3, 1987 (received for review May 28, 1987)*

**ABSTRACT** **Six nonallelic immunoglobulin λ constant region genes have been previously characterized on a 40-kilobase stretch of DNA. The nucleotide sequences of the three upstream genes of this cluster ($C_\lambda 1$, $C_\lambda 2$, $C_\lambda 3$) have been determined by other workers and shown to encode, respectively, the isotypic Mcg, Kern$^-$Oz$^-$, and Kern$^-$Oz$^+$ constant region of the λ chains. In this paper, we report the sequence of the three downstream genes of this cluster and show that two of them ($C_\lambda 4$ and $C_\lambda 5$) are pseudogenes. However, $C_\lambda 6$ encodes a Kern$^+$Oz$^-$ chain and corresponds to the fourth isotype described among the λ proteins sequenced so far. A potentially active $J_\lambda$ (joining) segment, with the canonical heptamer and nonamer sequences for rearrangement, is located 1.5 kilobases upstream of $C_\lambda 6$. The amino acid sequence encoded by the $C_\lambda 6$ gene is compared with the constant region sequences of various monoclonal Bence Jones λ proteins. Allotypic and isotypic differences confirm the polymorphism and complexity of the human $C_\lambda$ locus.**

In humans, the constant (C) region of the immunoglobulin λ light chains consists of at least four nonallelic or isotypic forms that differ by limited amino acid substitutions to produce the serological markers Kern (Ke) (1, 2), Oz (3–5), and Mcg (6, 7). Several additional substitutions have been described (8–16), but it is unknown whether these represent allelic variants or distinct isotypes. The human immunoglobulin λ light chain genes have been mapped to chromosome 22 (17) at band q11 (18, 19), and six nonallelic λ C region genes ($C_\lambda 1$ to $C_\lambda 6$) have been characterized on a 40-kilobase (kb) stretch of DNA (20). The number of $C_\lambda$ genes varies between six and nine per haploid genome (21). These variations were detected by restriction fragment length polymorphism (21) and seem to have arisen from unequal meiotic crossing-over with a duplication of the $C_\lambda 2$ and $C_\lambda 3$ genes. Moreover, three additional $C_\lambda$-like genes have been recently identified, which map on different stretches of DNA and are nonallelic (22). One of these is a pseudogene, whereas the two others encode a putative λ chain C region whose sequence differs from that of the λ chains described so far.

Only three $C_\lambda$ genes ($C_\lambda 1$, $C_\lambda 2$, and $C_\lambda 3$) belonging to the cluster described by Hieter have been sequenced (20), and they have been shown to encode, respectively, the Mcg, Ke$^-$Oz$^-$, and Ke$^-$Oz$^+$ C region of the λ chains. In this paper, we report the sequences* of the three genes located downstream in this cluster and show that two of them ($C_\lambda 4$ and $C_\lambda 5$) are pseudogenes, whereas $C_\lambda 6$ encodes a Ke$^+$Oz$^-$ chain, the fourth isotype described among the proteins sequenced so far. This $C_\lambda 6$ gene has a potentially active $J_\lambda 6$ joining region, with the canonical heptamer and nonamer sequences for rearrangement, 1.5 kb upstream of the coding C region.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

## MATERIALS AND METHODS

**Construction of a Phage Library from LY67 DNA.** DNA prepared from LY67 cells (a λ-producing Burkitt's lymphoma) (23) was partially digested with *Mbo* I. Restriction fragments 15–20 kb long were ligated into *Bam*HI-digested DNA of phage λ2001 (24) and packaged *in vitro*. Recombinant phages were screened by the *in situ* plaque hybridization procedure (25).

**Probes.** A genomic clone (Chr 22λ5) in λgt-λWES (26) was kindly provided by T. H. Rabbitts (Medical Research Council, Cambridge, England). This clone contains an 8.0-kb *Eco*RI fragment that includes the known nonallelic Ke$^-$Oz$^-$ ($C_\lambda 2$) and Ke$^-$Oz$^+$ ($C_\lambda 3$) genes and the flanking sequences (20). We subcloned a 700-base-pair (bp) *Bgl* II–*Eco*RI fragment containing only the Ke$^-$Oz$^-$ $C_\lambda 3$ gene (Fig. 1), and this $C_\lambda$ probe cross-hybridizes with all the other $C_\lambda$-like genes (20). It was radioactively labeled with [α-$^{32}$P]dCTP by nick-translation (27) and was used to screen the LY67 phage library.

**Subcloning and Sequencing Strategies.** One clone, LY67 $C_\lambda 3$-6 (Fig. 1), was shown to contain $C_\lambda 3$ to $C_\lambda 6$. Appropriate subclones were made in pUC vectors (29). Nucleotide sequence analysis was carried out by dideoxy chain-termination procedures (30) in M13 vectors (31) by deploying exonuclease III–nuclease S1 methods (32) or directed sequencing using known restriction enzyme sites.

**Oligonucleotide Synthesis and Hybridization.** A 19-mer oligonucleotide 5′ GTGTTCGGCGGAGGGACCA 3′ corresponding to part of the $J_\lambda 3$ gene segment sequence (this paper and ref. 28) was synthesized, radiolabeled, and hybridized to the LY67 $C_\lambda 3$-6 clone to search for other $J_\lambda$ segments. Low-stringency washes were carried out at room temperature.

## RESULTS

**Rearrangement of a $V_\lambda$III Subgroup Gene to $J_\lambda 3$ in the LY67 Cell Line.** One clone (LY67 $C_\lambda 3$-6) containing a 18-kb piece of genomic DNA was isolated and characterized. A restriction map of this clone is shown in Fig. 1. Comparison of this map with one previously published (20) suggested that this clone contains four $C_\lambda$ genes, namely $C_\lambda 3$ to $C_\lambda 6$. The sequence of the 5′ end of the LY67 $C_\lambda 3$-6 clone shows that a $V_\lambda$ gene rearrangement has occurred, joining this gene to the $J_\lambda 3$ gene segment, which is located 1.5 kb upstream of $C_\lambda 3$ (28). Fig. 2A shows the partial nucleotide sequence of the rearranged $V_\lambda$ in LY67 and that of a $V_\lambda$ gene assigned to the $V_\lambda$III subgroup and isolated from the Burkitt lymphoma cell line PA682 (28).

Abbreviations: C, constant; J, joining; V, variable; Ke, Kern.

*The sequences reported in this paper are being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) [accession nos. J03009 ($C_\lambda 4$), J03010 ($C_\lambda 5$), and J03011 ($C_\lambda 6$)].

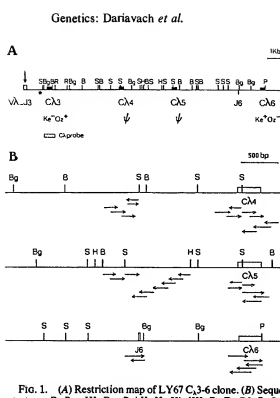



FIG. 1. (*A*) Restriction map of LY67 $C_\lambda 3$-6 clone. (*B*) Sequencing strategy. B, *Bam*HI; Bg, *Bgl* II; H, *Hind*III; R, *Eco*RI; S, *Sst* I. Of the *Pst* I sites (P), only the one used for subcloning the fragment containing the $C_\lambda 6$ gene is indicated. The rearrangement $V\text{-}J_\lambda 3$ is indicated by an arrow (*V*, variable region). An asterisk shows the location of a polymorphic *Bam*HI site present in PA682 DNA (28) but absent from our LY67 clone.

The deduced amino acid sequence of the rearranged *V* gene of LY67 is also compared to the V region of the protein DEL of the subgroup $V_\lambda$III (33, 34). A 75% sequence identity indicates that the $V_\lambda$ gene rearranged in LY67 is a member of the $V_\lambda$III subgroup gene family, and this is in agreement with the detection of a transcript hybridizing to a $V_\lambda$III probe in the LY67 cell line (28). The $J_\lambda 3$ segment of the LY67 $C_\lambda 3$-6 clone, compared to the $J_\lambda 3$ segment rearranged in PA682, shows two nucleotide differences (one of them resulting in a valine/leucine amino acid substitution) that may be due to allelic polymorphism. Two other nucleotide differences are observed at the *V–J* junction and are probably explained by a flexibility in the mechanism by which junctions occur (35, 36).

**$C_\lambda 6$ Encodes a $Ke^+Oz^-$ Chain.** Fig. 2*B* shows the nucleotide sequence of the $C_\lambda 6$ gene and the encoded amino acid sequence (106 residues). The residues Ala, Ser, and Thr, found, respectively, at codons 6, 8, and 57 (positions 112, 114, and 163 according to ref. 34) indicate that $C_\lambda 6$ encodes a $Mcg^-$ protein. Arg (codon 83, position 190) corresponds to the $Oz^-$ marker, whereas Gly (codon 46, position 152) characterizes the $Ke^+$ marker. Therefore the $C_\lambda 6$ gene encodes the fourth isotype $Ke^+Oz^-$.

**$J_\lambda 6$ Segment Is 1.5 kb Upstream of $C_\lambda 6$.** Only the $J_\lambda 1$ (22) and $J_\lambda 3$ segments (ref. 28 and this paper) have been characterized; they have been localized in genomic DNAs at 1.5 kb upstream of the respective $C_\lambda$ coding regions. We therefore used an oligonucleotide corresponding to the $J_\lambda 3$ sequence (see *Materials and Methods*) to search for homologous $J_\lambda$ segments in the LY67 $C_\lambda 3$-6 clone. As expected, a strong signal was obtained for the $J_\lambda 3$-containing fragments, whereas a weaker signal allowed us to detect the $J_\lambda 6$ segment in a *Sac* I–*Bgl* II fragment upstream of $C_\lambda 6$. The sequence of this $J_\lambda 6$ segment (Fig. 2*C*) showed that it encodes 12 amino acids (among them the characteristic Phe-Gly-Xaa-Gly residues) and that it also possesses the canonical heptamer and nonamer sequences essential to *V–J* rearrangement (37, 38). No signal corresponding to the putative $J_\lambda 4$ and $J_\lambda 5$ segments could be detected in the LY67 $C_\lambda 3$-6 clone by using either the oligonucleotides ($J_\lambda 3$ probe) or the genomic *Sac* I–*Bgl* II fragment ($J_\lambda 6$ probe), indicating that if these segments exist their homology is too weak to be detected in our conditions of hybridization. Since the $J_\lambda 3$ and $J_\lambda 6$ gene segments are 1.5 kb upstream of their respective coding regions, we subcloned fragments located, respectively, at about the same distance upstream of $C_\lambda 4$ and $C_\lambda 5$. Although in both cases, we detected some conserved heptamer sequences, we did not

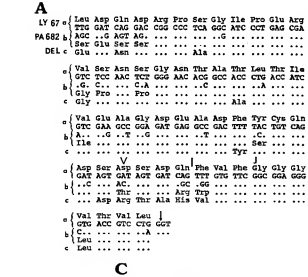



FIG. 2. Nucleotide and amino acid sequences. (*A*) Partial sequence of the LY67 $V_\lambda$–$J_\lambda 3$ rearranged gene. (*B*) Sequence of the $C_\lambda 6$ gene. (*C*) Sequence alignment of $J_\lambda 6$ and $J_\lambda 1$ (22).

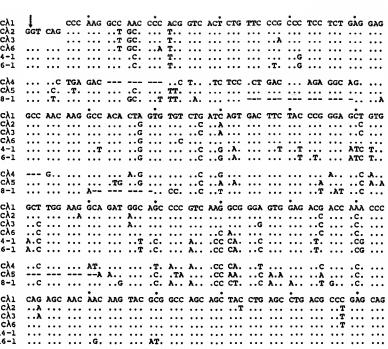


FIG. 3. Nucleotide sequence alignment of the human $C_\lambda$ genes. Dots indicate identity with $C_\lambda 1$; dashes indicate gaps.

find the characteristic Phe-Gly-Xaa-Gly residues or the expected splice site at a downstream position. It is possible that the heptamer sequences are attached to poorly conserved pseudo $J_\lambda$ segments and we cannot exclude the possibility that the putative $J_\lambda 4$ and $J_\lambda 5$ are localized in fragments that were not sequenced, upstream of the $C_\lambda 4$ and $C_\lambda 5$ genes.

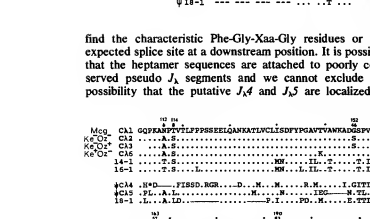


FIG. 4. Protein sequences derived from the human $C_\lambda$ gene sequences. The standard one-letter symbols are used.

***$C_\lambda 4$* and *$C_\lambda 5$* Are Pseudogenes.** Nucleotide sequences and the encoded amino acid sequences of $C_\lambda 4$ and $C_\lambda 5$ are shown in Figs. 3 and 4. Both genes are pseudogenes; the third codon of $C_\lambda 4$ is a stop codon, and $C_\lambda 4$ displays three deletions. The first deletion of 9 bp spans codons 5 to 7 and the other two deletions excise codons 21 and 64. $C_\lambda 5$ has an 11-bp deletion (codons 41–44) resulting in a frameshift.

Table 1. Amino acid differences between the four nonallelic forms of the human $C_\lambda$ regions

| Isotype | Gene | Amino acid residue 112 (6) | 114 (8) | 152 (46) | 163 (57) | 190 (83) |
|---|---|---|---|---|---|---|
| Mcg$^+$ Ke$^+$Oz$^-$ | $C_\lambda 1$ | Asn | Thr | Gly | Lys | Arg |
| Mcg$^-$ Ke$^-$Oz$^-$ | $C_\lambda 2$ | Ala | Ser | Ser | Thr | Arg |
| Mcg$^-$ Ke$^-$Oz$^+$ | $C_\lambda 3$ | Ala | Ser | Ser | Thr | Lys |
| Mcg$^-$ Ke$^+$Oz$^-$ | $C_\lambda 6$ | Ala | Ser | Gly | Lys | Arg |

Residue numbering is according to ref. 34; parentheses enclose numbering of the codons in the $C_\lambda$ genes.

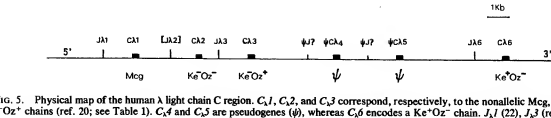



FIG. 5. Physical map of the human λ light chain C region. $C_\lambda 1$, $C_\lambda 2$, and $C_\lambda 3$ correspond, respectively, to the nonallelic Mcg, Ke$^-$Oz$^-$, and Ke$^-$Oz$^+$ chains (ref. 20; see Table 1). $C_\lambda 4$ and $C_\lambda 5$ are pseudogenes ($\psi$), whereas $C_\lambda 6$ encodes a Ke$^+$Oz$^-$ chain. $J_\lambda 1$ (22), $J_\lambda 3$ (ref. 28 and this paper), and $J_\lambda 6$ (this paper) have been localized 1.5 kb upstream of $C_\lambda 1$, $C_\lambda 3$, and $C_\lambda 6$, respectively. $J_\lambda 2$ has not yet been localized in genomic DNA. No $J_\lambda$ gene segment has so far been identified upstream of $C_\lambda 4$ and $C_\lambda 5$.

## DISCUSSION

In human λ chain C regions, the Mcg marker involves amino acid residues at positions 112, 114, and 163 (numbering according to ref. 34; Table 1) corresponding, respectively, to codons 6, 8, and 57 of the $C_\lambda$ genes. Mcg$^+$ proteins have residues Asn-112, Thr-114, and Lys-163, whereas Mcg$^-$ proteins have residues Ala, Ser, and Thr, respectively, at these locations. However, the recently sequenced Mor protein is different from the other Mcg$^-$ proteins by having Ala-163 instead of Thr-163 (16). The Ke and Oz markers occur at positions 152 and 190, respectively: Ke$^+$ proteins have Gly-152 and Ke$^-$ have Ser-152; Oz$^+$ proteins have Lys-190 and Oz$^-$, Arg-190. These markers define four nonallelic forms of human λ chain C regions (Mcg, Ke$^-$Oz$^-$, Ke$^-$Oz$^+$, and Ke$^+$Oz$^-$), which are encoded, respectively, by the $C_\lambda 1$, $C_\lambda 2$, $C_\lambda 3$ (20), and $C_\lambda 6$ (this paper) genes (Fig. 5).

In Fig. 6, monoclonal Bence Jones proteins have been assigned as products of the $C_\lambda$ genes 1, 2, 3, or 6 on the basis of the presence or absence of residues characteristic for the Mcg, Ke, and Oz markers. In most cases, there is complete concordance between the protein and the deduced amino acid sequence of the corresponding $C_\lambda$ gene. However, other amino acid changes have been found in several proteins (Table 2 and Fig. 6). Since these substitutions have been noted only once, they could represent allotypic differences. However, it is not excluded that some of the proteins Ke$^-$Oz$^-$ could be encoded by a $C_\lambda$ gene resulting from the duplication of the $C_\lambda 2$–$C_\lambda 3$ region, as has been described in some individuals (21). In such cases these sequences should represent new isotypic differences due to the presence of several nonallelic copies of $C_\lambda 2$ gene. Differences observed in the $J_\lambda 2$ sequences (Fig. 6) might for the same reason be either allotypic or isotypic. Differences in the $J_\lambda 1$ segment region might represent allotypic differences, although the presence of not yet identified other $J_\lambda 1$ segments cannot be ruled out.

If we compare the deduced amino acid sequence of $C_\lambda 6$ gene with two known Ke$^+$Oz$^-$ proteins that have identical $C_\lambda$ coding regions, SM (53) and Kern (52), the protein predicted for $C_\lambda 6$ shows three differences: (*i*) lysine at position 145 (codon 39) instead of threonine, (*ii*) asparagine at position 156 (codon 50) instead of lysine, and (*iii*) alanine at position 210

Table 2. Sequence variations in the C region of human λ chains

| Isotype | Common amino acid | Residue number | Variant amino acid | Protein | Ref. |
|---|---|---|---|---|---|
| Mcg | Lys | 156 | Glu | WEIR | 8 |
| Ke$^-$Oz$^-$ | Lys | 129 | Glx | CH | 9 |
| | | | Arg | NIG68 | 10 |
| | | | Ser | ATK | 11 |
| | Ala | 143 | Val | MZ | 12, 13 |
| | Val | 155 | Ile | HIL | 14 |
| | Lys | 156 | Glu | SA | 11 |
| | Ala | 157 | Val | WAY | 1 |
| | Lys | 172 | Asn | MZ | 12 |
| | | | Arg | MOR | 16 |
| | Lys | 187 | Gln | ATK | 11 |
| | Gln | 195 | Leu | NIG68 | 10 |
| Ke$^-$Oz$^+$ | Asp | 151 | Glu | EV | 15 |

Residue numbers according to Kabat *et al.* (34).

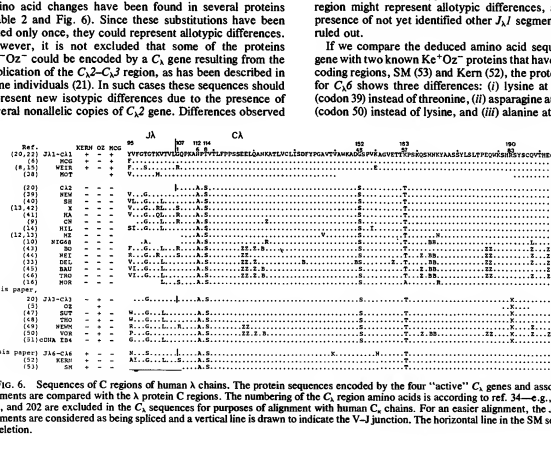



FIG. 6. Sequences of C regions of human λ chains. The protein sequences encoded by the four "active" $C_\lambda$ genes and associated $J_\lambda$ gene segments are compared with the λ protein C regions. The numbering of the $C_\lambda$ region amino acids is according to ref. 34—e.g., positions 169, 201, and 202 are excluded in the $C_\lambda$ sequences for purposes of alignment with human $C_\kappa$ chains. For an easier alignment, the $J_\lambda$ and $C_\lambda$ gene segments are considered as being spliced and a vertical line is drawn to indicate the V–J junction. The horizontal line in the SM sequence shows a deletion.

(codon 103) instead of threonine (Fig. 6). These differences may represent allotypic variations, although we cannot entirely exclude the possibility that these different $Ke^+Oz^-$ sequences are encoded by nonallelic genes. More sequences of $C_\lambda$ genes or λ proteins should help estimate the extent of the human λ chain polymorphism.

Note Added in Proof. The $J_\lambda 2$ gene segment has recently been localized upstream of $C_\lambda 2$ (54).

We are very grateful to Dr. T. H. Rabbitts for providing us with the LY67 cell line and Chr22λ5 clone. We thank Dr. P. Fort for his collaboration, which was greatly appreciated. We also thank C. Aymard for synthesizing the oligonucleotide and Véronique Ramond for typing the manuscript. This work was supported by the Centre National de la Recherche Scientifique, the Association pour la Recherche sur le Cancer, the Fondation pour la Recherche Médicale, and the Ministère de la Recherche.

1. Ponstingl, H., Hess, M. & Hilschmann, N. (1968) *Hoppe-Seyler's Z. Physiol. Chem.* **349**, 867–871.
2. Hess, M., Hilschmann, N., Rivat, L., Rivat, C. & Ropartz, C. (1971) *Nature (London) New Biol.* **234**, 58–61.
3. Ein, D. & Fahey, J. L. (1967) *Science* **156**, 947–950.
4. Ein, D. (1968) *Proc. Natl. Acad. Sci. USA* **60**, 982–985.
5. Appella, E. & Ein, D. (1967) *Proc. Natl. Acad. Sci. USA* **57**, 1449–1454.
6. Fett, J. W. & Deutsch, H. F. (1974) *Biochemistry* **13**, 4102–4114.
7. Fett, J. W. & Deutsch, H. F. (1975) *Immunochemistry* **12**, 643–652.
8. Jabusch, J. R. & Deutsch, H. F. (1982) *Mol. Immunol.* **19**, 901–906.
9. Okada, Y., Nozu, Y., Titani, K., Watanabe, S., Hara, H. & Kitagawa, M. (1972) *Immunochemistry* **9**, 207–210.
10. Shinoda, T., Yoshimura, K., Kametani, F. & Isobe, T. (1983) *Biochem. Biophys. Res. Commun.* **117**, 587–592.
11. Lieu, T. S., Deutsch, H. F. & Tischendoif, F. W. (1977) *Immunochemistry* **14**, 429–433.
12. Milstein, C. (1967) *Nature (London)* **216**, 330–332.
13. Milstein, C., Frangione, B. & Pink, J. R. L. (1967) *Cold Spring Harbor Symp. Quant. Biol.* **32**, 31–36.
14. Lopez de Castro, J. A., Chiu, Y.-Y. H. & Poljak, R. J. (1978) *Biochemistry* **17**, 1718–1723.
15. Fett, J. W. & Deutsch, H. F. (1976) *Immunochemistry* **13**, 149–155.
16. Frangione, B., Moloshok, T., Prelli, F. & Solomon, A. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 3415–3419.
17. Erickson, J., Martinis, J. & Croce, C. M. (1981) *Nature (London)* **294**, 173–175.
18. McBride, O. W., Hieter, P. A., Hollis, G. F., Swan, D., Otey, M. C. & Leder, P. (1982) *J. Exp. Med.* **155**, 1480–1490.
19. de la Chapelle, A., Lenoir, G., Boué, J., Gallano, P., Huerre, C., Szajnert, M. F., Jeanpierre, M., Lalouel, J. M. & Kaplan, J. C. (1983) *Nucleic Acids Res.* **11**, 1133–1142.
20. Hieter, P. A., Hollis, G. F., Korsmeyer, S. J., Waldmann, T. A. & Leder, P. (1981) *Nature (London)* **294**, 536–540.
21. Taub, R. A., Hollis, G. F., Hieter, P. A., Korsmeyer, S., Waldmann, T. A. & Leder, P. (1983) *Nature (London)* **304**, 172–174.
22. Chang, H., Dmitrovsky, E., Hieter, P. A., Mitchell, K., Leder, P., Turoczi, L., Kirsch, I. R. & Hollis, G. F. (1986) *J. Exp. Med.* **163**, 425–435.
23. Lenoir, G. M., Preud'homme, J. L., Bernheim, A. & Berger, R. (1982) *Nature (London)* **298**, 474–476.
24. Karn, J., Matthes, H. W. D., Gait, M. J. & Brenner, S. (1984) *Gene* **32**, 219–224.
25. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
26. Rabbitts, T. H., Forster, A. & Matthews, J. (1983) *Mol. Biol. Med.* **1**, 11–19.
27. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
28. Sun, L.-H. K., Croce, C. M. & Showe, L. C. (1985) *Nucleic Acids Res.* **13**, 4921–4934.
29. Vieira, J. & Messing, J. (1982) *Gene* **19**, 259–268.
30. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
31. Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H. & Roe, B. A. (1980) *J. Mol. Biol.* **143**, 161–178.
32. Guo, L. H., Yang, R. C. A. & Wu, R. (1983) *Nucleic Acids Res.* **11**, 5521–5540.
33. Eulitz, M. (1974) *Eur. J. Biochem.* **50**, 49–69.
34. Kabat, E. A., Wu, T. T., Bilofsky, H., Miller-Reid, M. & Perry, H. (1983) *Sequences of Proteins of Immunological Interest* (Natl. Inst. of Health, Washington, DC).
35. Max, E. E., Seidman, J. G. & Leder, P. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 3450–3454.
36. Sakano, H., Hüppi, K., Heinrich, G. & Tonegawa, S. (1979) *Nature (London)* **280**, 288–294.
37. Early, P., Huang, H., Davis, M., Calame, K. & Hood, L. (1980) *Cell* **19**, 981–992.
38. Kojima, M., Odani, S. & Ikenaka, T. (1980) *Mol. Immunol.* **17**, 1407–1414.
39. Langer, B., Steinmetz-Kayne, M. & Hilschmann, N. (1968) *Hoppe-Seyler's Z. Physiol. Chem.* **349**, 945–951.
40. Titani, K., Wikler, M., Shinoda, T. & Putnam, F. W. (1970) *J. Biol. Chem.* **245**, 2171–2176.
41. Shinoda, S., Titani, K. & Putnam, F. W. (1970) *J. Biol. Chem.* **245**, 4475–4487.
42. Milstein, C., Clegg, J. B. & Jarvis, J. M. (1968) *Biochem. J.* **110**, 631–652.
43. Wikler, M. & Putnam, F. W. (1970) *J. Biol. Chem.* **245**, 4488–4507.
44. Garver, F. A. & Hilschmann, N. (1972) *Eur. J. Biochem.* **26**, 10–32.
45. Baczko, K., Braun, D. G., Hess, M. & Hilschmann, N. (1970) *Hoppe-Seyler's Z. Physiol. Chem.* **351**, 763–767.
46. Scholz, R. & Hilschmann, N. (1975) *Hoppe-Seyler's Z. Physiol. Chem.* **356**, 1333–1335.
47. Frangione, B., Moloshok, T. & Solomon, A. (1983) *J. Immunol.* **131**, 2490–2493.
48. Ohiso, J., Solomon, A. & Frangione, B. (1986) *J. Immunol.* **136**, 716–719.
49. Chen, B. L. & Poljak, R. J. (1974) *Biochemistry* **13**, 1295–1302.
50. Engelhard, M., Hess, M. & Hilschmann, N. (1974) *Hoppe-Seyler's Z. Physiol. Chem.* **355**, 85–88.
51. Anderson, M. L. M., Brown, L., McKenzie, E., Kellow, J. E. & Young, B. D. (1985) *Nucleic Acids Res.* **13**, 2931–2941.
52. Ponstingl, H., Hess, M. & Hilschmann, N. (1971) *Hoppe-Seyler's Z. Physiol. Chem.* **352**, 247–266.
53. Garver, F. A., Chang, L., Mendicino, J., Isobe, T. & Osserman, E. F. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 4559–4563.
54. Udey, J. A. & Blomberg, B. (1987) *Immunogenetics* **25**, 63–70.

The EMBO Journal vol.9 no.5 pp.1535–1542, 1990

# Transposition of human immunoglobulin $V_x$ genes within the same chromosome and the mechanism of their amplification

F.-J.Zimmer, H.Hameister[1], H.Schek and H.G.Zachau

Institut für Physiologische Chemie, Physikalische Biochemie und Zellbiologie der Universität München, 8000 München 2 and [1]Abteilung Klinische Genetik der Universität Ulm, 7900 Ulm, FRG

Communicated by H.G.Zachau

**The variable, joining and constant gene segments of the human immunoglobulin $\kappa$ locus ($V_\kappa$, $J_\kappa$ and $C_\kappa$) are located on the short arm of chromosome 2 at 2p11–2p12. Here we describe a cluster of 11 $V_\kappa$ genes on the long arm of chromosome 2 at 2cen–q11. By pulsed-field gel electrophoresis, cosmid cloning and DNA sequencing the cluster was shown to consist of four amplified units (amplicons). The amplicons, each 110–160 kb in size, are organized within 650 kb as an array of inverted repeats with short stretches of non-amplified DNA in between. Cloning and sequencing of three different joints between amplified and non-amplified DNA revealed the existence of parts of Alu repeats at each of the analysed joints. It is suggested that during evolution a group of five $V_\kappa$ genes was transposed from the short to the long arm of chromosome 2 by a pericentric inversion. Three of the five $V_\kappa$ genes were then amplified in two subsequent steps to yield the structure found in the majority of the present day population. The possible relation of this structure to a pericentric inversion of chromosome 2 that is seen cytogenetically in a small fraction of today's population is discussed.**

*Key words:* Alu repeats/amplification/immunoglobulin $V_\kappa$ genes/transposition

## Introduction

Amplification and transposition of genes play an important role in the formation of multigene families during evolution (for review, see Maeda and Smithies, 1986). In the case of the human gene family coding for the variable regions of the immunoglobulin light chains of the $\kappa$ type ($V_\kappa$), putative transpositions of $V_\kappa$ genes led to the formation of a 'mixed' gene cluster, in which genes of different subgroups are interdigitated (Pech and Zachau, 1984). A subsequent duplication led to the generation of a second copy of a large part of the $V_\kappa$ gene cluster (for reviews, see Zachau, 1989, 1990). The $\kappa$ locus contains ~70 $V_\kappa$ genes (for reviews, see Zachau, 1989, 1990) and is located on the short arm of chromosome 2, at 2p12 (Malcolm *et al.*, 1982). In addition, $V_\kappa$ genes have been found on the chromosomes 1, 22 (Lötscher *et al.*, 1986, 1988) and other chromosomes (Straubinger *et al.*, 1988). These genes are called orphons is analogy to histone and ribosomal RNA genes found outside of their respective gene clusters (Childs *et al.*, 1981).

We have previously reported the structure of two contiguous cloned regions (contigs), called Wa and Wb, with a total of nine $V_\kappa$ genes (Pohlenz *et al.*, 1987). The two W contigs have been assigned to chromosome 2 (Lötscher *et al.*, 1988) and were thought to be part of the $\kappa$ locus. However, we did not succeed in linking them to the cloned parts of the $\kappa$ locus by chromosomal walking or by pulsed-field gel electrophoresis (PFG). In the course of the PFG experiments, a third W contig was detected. The three contigs are present in the genomes of all individuals so far analysed. The observation that the W contigs are located on chromosome 2, yet are not part of the $\kappa$ locus, prompted us to analyse their genomic organization and chromosomal location in more detail.

## Results

### *Characterization of the W contigs*

We previously described the characterization of the two contigs Wa and Wb (Pohlenz *et al.*, 1987). Here, we report the isolation and characterization of two sets of cosmid clones, one extending the contig Wb and one representing a new W contig, termed Wc.

In the course of chromosomal walking experiments, the cosmid libraries III (Pohlenz *et al.*, 1987) and V (Lorenz, 1989) were screened with the W-specific clone m654-1 (Pohlenz *et al.*, 1987); 25 cosmid clones were isolated. Five of them were derived from Wa without extending the known contig. The other 20 clones belonged to Wb with some of them extending the contig. Two additional $V_\kappa$ genes were found in the extending cosmids. A map of the W contigs with some representative cosmid clones is shown in Figure 1. A description of all clones can be found in Zimmer (1989).

The existence of a third contig, Wc, was demonstrated on PFG blots that had been hybridized with the Wa-derived probe m167-1 (Pohlenz *et al.*, 1987; see also Figure 1). In subsequent restriction mapping of cosmid clones that had been isolated previously with m167-1, but not fully characterized (Pohlenz, 1986), three clones matching neither Wa nor Wb were identified. Wc was shown to be a third independent contig and not an allelic variant of Wa or Wb by demonstrating the existence of Wa-, Wb- and Wc-characteristic fragments in the genomes of all 20 unrelated individuals tested (Zimmer, 1989; data not shown here). The map of Wc is included in Figure 1.

We use the term amplified unit or amplicon for those parts of the contigs Wa, Wb and Wc, that are homologous to each other. As Wb contains two amplified units, a minimum of four W amplicons (named I–IV; Figure 1) exist within the human genome.

### *How many $V_\kappa$ gene containing W amplicons exist in the human genome?*

To test whether more than three $V_\kappa$ gene-containing amplicons exist, we estimated their copy number relative to $C_\kappa$, which is known to be single copy, by quantitation

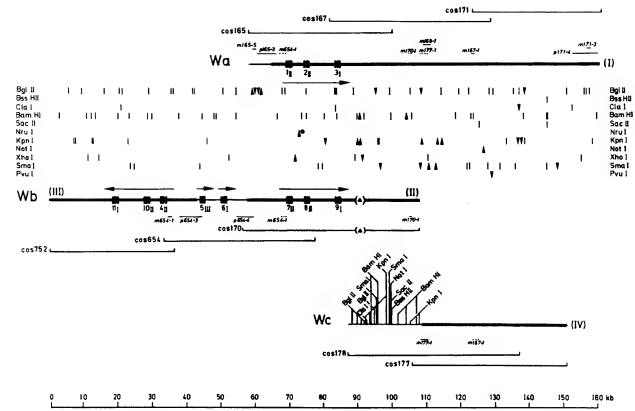



**Fig. 1.** Restriction maps of the genomic contigs Wa, Wb and Wc. The maps were derived from cosmid clones described in Pohlenz *et al.* (1987) and Zimmer (1989). For each contig only some representative cosmid clones are shown. The maps are aligned to demonstrate the homology between the contigs. Positions of restriction sites that are identical in two contigs or occur in contigs without counterpart are marked by vertical bars. Differences in restriction sites between two contigs are symbolized by arrows pointing to the respective map positions. In the map of Wc, restriction sites are given separately for the part that shows no homology to Wb or Wa. The scale (kb) applies to all contigs. A deletion of 1 kb in Wb at map position 92 kb is marked by a triangle. $V_x$ genes are drawn as filled boxes; the subgroup designations are indicated (I–III). Plasmid and M13 subclones are shown as horizontal bars above the restriction map of Wa and underneath those of Wb and Wc. The subclones m654-1, m165-5 and m167-1 are described in Pohlenz *et al.* (1987), the remaining ones in Zimmer (1989). Amplified units are marked by fat lines and are numbered (I–IV). Clones derived from one amplicon hybridize to the corresponding position within other amplicons; these positions are marked by dotted lines. Arrows at the maps of Wa and Wb indicate the transcriptional orientation of the $V_x$ genes. A dot marks a *Nru*I site at map position 72 kb, which is present only in some alleles of Wa.

of Southern blot hybridizations. We chose a clone which hybridizes to a position close to one of the amplified $V_x$ genes (m654-1) and linked it to a fragment derived from the $C_x$ region (I-1). The construct m654-1/I-1 is shown in Figure 2a and the blot hybridizations are in Figure 2b. The copy number of the W contigs is estimated as the multiple of the $C_x$ signal (Table I). From the calculated W/$C_x$ ratios it is very likely that no further amplicons hybridizing with m645-1 exist. The procedure of copy number determination is similar to the one of Meindl (1990) who estimated the overall number of $V_x$ genes including W-type $V_x$ genes in the DNA of individuum AF, and arrived at data compatible with three copies of the W amplicon and one copy of the non-amplified gene W6. The cosmid clones constituting the W contigs of Figure 1 are derived from the DNA of individuum N and individuum St. Since the blot hybridizations of restriction nuclease digests of the DNAs AF, N, St and PC-3 (see below) show no significant differences in the regions of the W-type $V_x$ genes (Meindl, 1990) it is likely that the number of $V_x$ gene-containing W amplicons is three. The amplified unit within Wc is not detected by hybridization with m654-1/I-1 since it does not reach into the $V_x$ gene-containing regions of the W amplicons (Figure 1). To identify other possibly existing truncated amplicons and to elucidate the genomic organization of the known amplicons, we established a long-range map of all regions hybridizing with W-specific probes by PFG.

### *Long-range map of the W regions*

The analysis of the organization of the W region was carried out with DNA from the prostate carcinoma cell line PC-3 (Kaighn *et al.*, 1979), as PC-3 DNA seems to contain more unmethylated restriction sites than DNA from other sources (Lorenz *et al.*, 1987), allowing the use of methylation sensitive restriction nucleases for long-range mapping.

A remarkable feature of the PFG blots is that all W region probes tested hybridize only to two *Not*I fragments, 250 and 600 kb in size. $V_x$ gene probes hybridize mainly with two additional *Not*I fragments which contain the *x* locus (Lorenz *et al.*, 1987). The identification of *Not*I sites in cosmid clones of Wa and Wc, which are also present in PC-3 DNA (*Not*I sites at map positions 550 and 1150 kb of Figure 3), has been useful in the construction of the long-range map.

The PFG studies did not reveal any evidence for the existence of additional W regions. According to the PFG experiments, the four W amplicons are organized as an array of large inverted repeats.

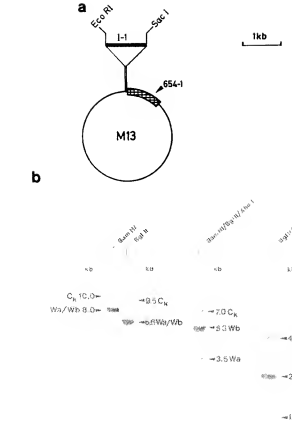



**Fig. 2.** Determination of the number of W contigs. In (**a**) the construct m654-1/1-1, which was used for the hybridizations, is shown. A 1.1 kb *Eco*RI–*Sac*I fragment from pI-1, which is unique for the $C_x$ region (Klobeck *et al.*, 1984), has been cloned into the W-specific probe m654-1. Southern blot analyses with m654-1/I-1 are shown in (**b**). Each lane contains 10 μg of placenta DNA digested with the indicated restriction nucleases. The W and $C_x$ derived fragments are assigned to the respective bands, the sizes of which are given in kb. For the quantitative evaluation see Table I.

**Table I.** Determination of the number of $V_x$ gene-containing W amplicons

| Restriction digests | Size of bands[a] (kb) | Region[a] | Counts/min[b] | Copy number[c] |
|---|---|---|---|---|
| *Bam*HI | 10.0 | $C_x$ | 14 ± 10 | 1.0 |
| | 8.0 | Wa+Wb | 47 ± 9 | 3.5 (3–4) |
| *Bgl*II | 9.5 | $C_x$ | 21 ± 10 | 1.0 |
| | 6.6 | Wa+Wb | 58 ± 9 | 2.8 (3) |
| *Bam*HI/*Bgl*II/ *Xho*I | 7.0 | $C_x$ | 43 ± 7 | 1.0 |
| | 5.3 | Wb | 87 ± 7 | 2.0 (2) |
| | 3.5 | Wa | 32 ± 8 | 0.7 (1) |
| *Bgl*II/*Sph*I | 4.8 | $C_x$ | 40 ± 7 | 1.0 |
| | 2.5 | Wb | 80 ± 7 | 2.0 (2) |
| | 1.5 | Wa | 35 ± 8 | 0.9 (1) |

[a]The assignment of bands in the digests to certain contigs is seen in Figure 2b.
[b]The values are corrected for blank values as described in Materials and methods. The standard deviations are indicated.
[c]The copy number for $C_x$ is taken as 1.0. The W/$C_x$ ratios are calculated by dividing the c.p.m. of the W containing bands by the one for $C_x$; the resulting numbers of W copies are given in parentheses.

### *Cloning and sequence analyses of joints between amplified and non-amplified DNA*

Four joints between amplified and non-amplified DNA can be identified within the W regions by comparing the respective restriction maps (transitions between bars and lines in Figure 1). By cloning joint-containing fragments into plasmid (p165-3, p654-3, p654-4) or M13 vectors (m177-1, m168-1) and comparing the fine maps, fragments suitable for the sequence analyses of the joints were identified and subcloned.

The break-off in homology between Wa and Wb was localized by aligning the map of Wa with that of the 5′ amplified unit of Wb (amplicon III) and comparing the restriction maps of p165-3 and p654-3. The joints of the amplicons II and III in Wb were localized in a similar manner by comparing the maps of p654-3 and p654-4. To localize the Wc joint, the maps of m168-1 and m177-1 were aligned (Figure 1). The sequences of the analysed joints are shown in Figure 4. It is evident from these data that Alu repeats played a role in the recombination processes, which led to the formation of the novel joints.

### *The W contigs reside on the long arm of chromosome 2*

As we have not been able to link by chromosomal walking or PFG the W contigs to the *x* locus, we tried to determine the chromosomal location of W by analysing somatic cell hybrids.

In one set of experiments, mouse–human hybrid cell lines, which contain only parts of human chromosome 2, were analysed with probes specific for the *x* locus or the W regions. In one of the lines, JI4-2L (Erikson *et al.*, 1983), only that part of the short arm of chromosome 2 is present that comprises $C_x$ to the telomere (2p12–tel). DNA of this cell line hybridizes only with a $C_x$-specific probe but with neither probes derived from the $V_x$ gene-containing parts of the locus nor the W region-specific probe m654-1 (Zimmer, 1989).

A second analysed cell line, RRP5-3 (Shiloh *et al.*, 1985), represents the almost complementary situation to JI4-2L, as RRP5-3 contains a $2p^-$ chromosome, i.e. the long arm and only that part of the short arm between the centromere and the *x* locus in 2p12. DNA of this cell line does not hybridize with any of the single-copy probes specific for the *x* locus; however, it gives a strong signal with m654-1 (Zimmer, 1989). According to these data, the W contigs are located either on the short arm of chromosome 2, between the *x* locus and the centromere, or on the long arm.

A more precise localization of the W regions was achieved by *in situ* hybridizations; these results are shown in Figure 5. The *x* locus specific probe pC-2 (Klobeck *et al.*, 1984) maps to the short arm of chromosome 2 in the region 2p11–2p12 in accordance with the location of the *x* locus (Malcolm *et al.*, 1982; McBride *et al.*, 1982). The W-derived cosmid clones cos178, cos165 (Figure 1) and cos140 (Pohlenz *et al.*, 1987) as well as the subclone m654-1 (data not shown) map to a region close to the centromere on the long arm of

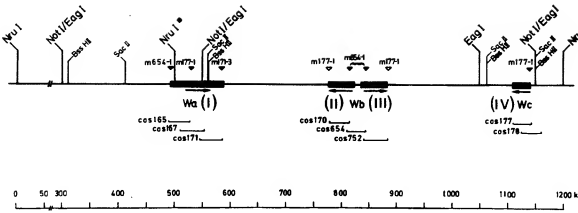



Fig. 3. Long range map of the W contigs. The map was constructed on the basis of numerous PFG blots, most of which were hybridized consecutively with different probes in order to find out whether certain fragments are recognized by more than one probe [a list of fragments hybridizing with W-specific probes can be found in Zimmer (1989)]. The shown orientation of Wb is arbitrary, as there is no cloned restriction site within Wb which would allow one to determine its orientation. Wa and Wc were oriented by identification of restriction sites for *Not*I and *Nru*I in cosmids of the Wa contig, and of a *Not*I site in cosmids of the Wc contig. These sites are also present in DNA of PC-3 which is used in the PFG experiments (map positions 500, 550 and 1150 kb). The distance between the Wa and Wc contigs is defined by the 600 kb *Not*I fragment which hybridized with m171-3 and m177-1. The contig Wb is placed in the centre of this fragment since we assume about equal sizes of the amplicons (see Discussion). The 500 and 700 kb *Nru*I fragments are defined by hydridization with m654-1, m171-3 and m177-1. The amplicons I–IV are shown as black bars. Arrows indicate the amplicon orientation based on the transcriptional orientation of $V_\lambda$ genes (I–III); the orientation of amplicon IV is based on its restriction map which is homologous to that of amplicon 1 (Figure 1). Filled and open triangles mark the regions to which the indicated probes hybridize. Open symbols indicate that the hybridizing region is not represented by cosmid clones. The terminal part of the map without known restriction sites is shortened (–//–). The analysed cell line PC-3 (Kaighn *et al.*, 1979) is heterozygous for the *Nru*I site marked by a rhombus. The linking of Wa to Wb and Wb to Wc is based on those restriction fragments marked with larger letters [for details see Zimmer (1989)].

chromosome 2 (2cen–q11). This position is obviously distinct from the *x* locus.

## Discussion

### *Size, copy number and organization of the W region amplicons*

According to the PFG experiments and the copy number determinations, four amplified units hybridizing with W region-specific probes reside within a DNA stretch of 650 kb and are organized as an array of inverted repeats. The cloned parts of the amplicons I–IV are 50–100 kb each (Figure 1). If the length of non-amplified DNA between the amplicons I and II and between III and IV (Figure 3) is similar to that between the amplicons II and III, the average size of the amplicons is in the range 110–160 kb. The chromosomal organization of the W amplicons, i.e. variable amplicon size, head-to-head or tail-to-tail orientation of the amplified units and stretches of non-amplified DNA in between two amplicons, is very similar to the organization of, for example, the amplified genes of dihydrofolate reductase (DHFR), CAD, c-*myc* and adenylate deaminase (AMPD) (DHFR, Ma *et al.*, 1988; Heartlein and Latt, 1989; CAD, Ardeshir *et al.*, 1983; Ford and Fried, 1986; c-*myc*, Ford and Fried, 1986; AMPD, Hyrien *et al.*, 1988).

One major difference between the organization of the W regions and the amplified units reported in the literature is the low copy number and a rather small amplicon size. The amplicons described so far exist either in a low copy number, which is the case after first step selection, but have a size up to 10 000 kb, found for CAD amplicons (Giulotto *et al.*, 1986) or the size is in the range of that of the W region amplicons, but the copy number reaches a value up to 2600 copies per cell, as has been described for AMPD amplicons (Yeung *et al.*, 1983). Whether these differences reflect a different mechanism of amplification responsible for the generation of the W regions is discussed in a following section.

### *Novel joints are formed within Alu repeats*

Relatively few studies define the DNA sequences at the sites of recombination associated with amplification (for review, see Stark *et al.*, 1989). The sequence analyses of novel joints revealed the existence of partial Alu repeats. This finding makes it very likely that the repeats played a role either during the amplification process itself or in recombinations, which are necessary to resolve aberrant replication bubbles into a linear array of amplified units. Depending on the assumed mechanism of amplification (for review, see Stark *et al.*, 1989), one can imagine several ways in which Alu repeats participate in the amplification process. In the context of strand switch models (Nalbantoglu and Meuth, 1986; Hyrien *et al.*, 1988), Alu repeats could serve as those sequences which promote the strand switch by the DNA polymerase within the replication fork. According to such a model, one would expect to find partial Alu repeats at the amplification joints if the strand switch event did not occur at identical positions on the leading and lagging strand.

In the course of recombination, which is always associated with DNA amplification, Alu repeats might serve as target sites for the recombinations. The involvement of Alu repeats in genomic recombinations has been reported frequently (see literature cited in Hyrien *et al.*, 1987).

Although we cannot decide at which step during the generation of the W amplicons the Alu repeats played a role, these repetitive sequence elements seemed to be important for the amplification of a W region precursor.

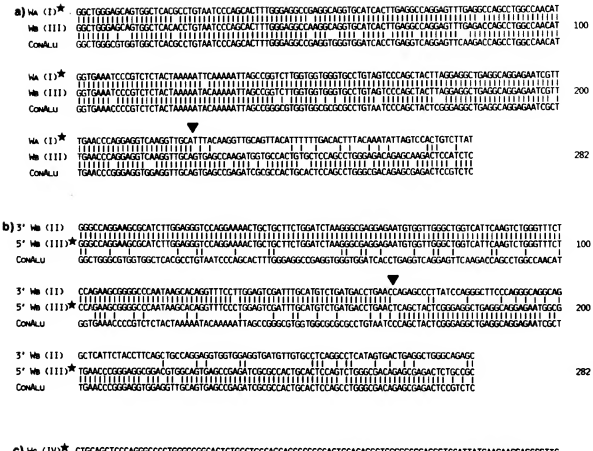



Fig. 4. Sequences of fragments, spanning joints between amplified and non-amplified DNA. The sequences are aligned to show maximal homology. Nucleotide positions which are identical between two sequences are marked by vertical bars. The regions from which the sequences are derived are indicated. ConAlu is a consensus sequence of human Alu repeats (Kariya *et al.*, 1987). The break-off in homology between two sequences is marked by a filled triangle. Those sequences which are believed to be the result of the recombinations associated with the generation of respective amplicons ('recombined sequences') are marked by asterisks. The non-marked sequences have to be considered as 'reference sequences'. which allow detection of the nucleotide position in the recombined sequence at which amplified and non-amplified DNA is joined. The sequences including some extensions that are not shown here are being transmitted to the EMBL Data Library. The sequencing strategies are described in Zimmer (1989). Both strands were sequenced in all cases. The sequence Wa in (a) is derived from a subcloned fragment of p165-3, spanning the joint between amplicon I and non-amplified DNA; the 'reference sequence' Wb is derived from a subcloned fragment of p654-3 (Figure 1). In (b) the sequence 5' Wb is derived from a fragment that contains the border of amplicon III; it was subcloned from p654-3 (Figure 1; Zimmer, 1989). The reference sequence, 3' Wb, is derived from a fragment that contains the border of amplicon II; the fragment was subcloned from p654-4 (Figure 1). The sequence Wc in (c) is derived from a subcloned fragment of m177-1 which contains the break off in homology to m168-1 (Figure 1). The boxed block of 56 nucleotides in Wa shows 90% sequence identity to the consensus sequence of human Alu repeats; the 5' part of the complementary strand is shown. The Alu sequence ends at position 177. The sequence 3' of position 177 in Wa shows 80% sequence identity to human LINE repeats (Skowronski *et al.*, 1988).

### *The W regions: products of two successive amplifications*

The results of sequence comparisons of genes derived from amplicons I–III make it very likely that these three amplicons were generated at different times and, hence, in a multistep process during evolution (Zimmer *et al.*, 1990). According to these data, the amplicons II and III, which form the large inverted repeat within Wb (Figure 1), were formed later in evolution than amplicon I. This conclusion is also supported by the results of restriction map comparisons of

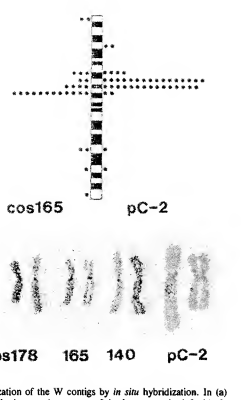



**Fig. 5.** Localization of the W contigs by *in situ* hybridization. In (a) the grain distribution on chromosome 2 is shown; on the left side for the W-derived cosmid clone cos165 and on the right side for the $C_x$ probe pC-2. For the probes cos178, cos140 and m654-1 the same localization was found as shown for cos165. In (b) arranged pairs of chromosome 2 are shown after hybridization with the indicated probes. Cosmid probes were detected by a non-radioactive procedure, pC-2 and m654-1 after radioactive labelling.

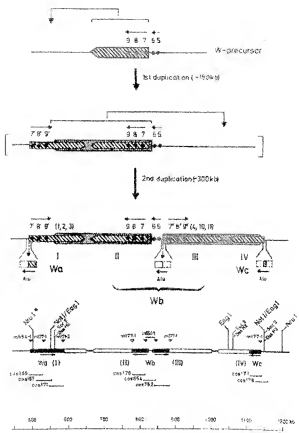



**Fig. 6.** Model of the generation of the W regions in a multistep process. The precursor contained five genes, marked by filled dots. Arrows indicate transcriptional orientations of genes. The section which becomes duplicated in the course of the first amplification event and the generated amplicons are drawn as hatched arrows with the arrow-heads pointing to the 3′ end with respect to the orientation of the genes. The brackets indicate that this structure does not exist in the genome anymore. That part of the intermediate product which is duplicated in a second event and the resulting amplicons are marked by dotted arrows. The four generated amplicons are termed I–IV in accordance with Figures I and 3. Genes generated in the first step are marked by ′, those of the second step by ″. The gene numbers of the present day W regions are given in parentheses. Parts of Alu repeats that have been detected at the joints are indicated. At the amplicon I joint, the 5′ part of an Alu repeat is part of the amplicon (hatched rectangle), the deleted 3′ part is marked by an open rectangle with a dotted line. At the joint of amplicon III, the 3′ part of an Alu repeat is present (open rectangle); it is part of the non-amplified DNA 5′ of amplicon III. At the amplicon IV joint only a few nucleotides of an Alu repeat are present (Figure 4c). The deleted 5′, as well as 3′, sequences are marked by open rectangles with dotted lines. The lower panel shows a long-range map of the W regions, similar to that shown in Figure 3. The open parts of the arrows, which indicate amplicons I–IV, represent those amplified DNA stretches that have not yet been cloned.

amplicons I–III, which show that of 29 sites mapped within a stretch of 40 kb only three sites differ between amplicons II and III. It is very likely that the three differences were created by a single event only, which led to a deletion of 1 kb in amplicon II (Figure 1). More differences are found between amplicons II and I (10 out of 33 sites differ) and between III and I (seven out of 33 sites differ). The evolutionary relationship of amplicon IV to the other three amplicons is hard to prove with this strategy, as those parts of amplicons II and III that are homologous to amplicon IV have not yet been cloned (Figure 1).

A chain of events which could have led to the generation of the W regions during evolution is shown in Figure 6. One assumption of the scheme is that the W region precursor had a structure similar to that part of Wb which contains the genes W5 to W9. This is reasonable to assume, as the five genes have the same transcriptional orientation (Figure 1), which also holds true for most genes of the *x* locus, from where the W precursor must be derived. In a first step, a section of ~150 kb with the genes W7, W8 and W9 is duplicated in such a way that a large palindrome is formed. The newly generated genes 7′, 8′ and 9′ represent the genes W1, W2 and W3 of Wa. In a second step, which led to the duplication of a large block of DNA of ~300 kb, a large palindrome is again formed. A third copy of the genes 7, 8 and 9 (7″, 8″ and 9″) is generated; these copies represent W4, W10 and W11 of Wb. Amplicon IV is a copy of the 3′ part of amplicon I, which has been formed in the first step.

A remarkable feature of this series of amplifications is that in the first, as well as in the second, step only one additional copy is formed. This is in contrast to the generation of other amplified units, where the first step of selection yields only a few additional copies; second step selections, and, hence, secondary amplifications, however, usually result in high copy numbers (Saito *et al.*, 1989). Alternatively, the low W copy numbers can be explained by secondary deletions of most of the W copies formed in the amplification events.

### *Transposition and amplification of the W regions*

For the generation of the W regions we propose the following chain of events: (i) the W region precursor, containing the genes W5 to W9 (Figure 6), was transposed to the long arm of chromosome 2 and (ii) the new chromosomal location of the precursor somehow promoted its stepwise amplification.

A pericentric inversion, involving chromosomal bands 2q1 and 2p1, could have been responsible for the transposition of the five V$_\lambda$ genes from the *x* locus on the short arm to the long arm of the chromosome. Such a chromosomal rearrangement is observed at the evolutionary progenitor of the chimpanzee (Yunis and Prakash, 1982). Transposition of V$_\lambda$ genes by such a process is consistent with a finding by Graninger *et al.* (1988), who detected another part of the *x* locus, a copy of the *x*-deleting element (*x*de; Siminovitch *et al.*, 1985), on the long arm of chromosome 2 in 2q11 (Graninger *et al.*, 1988). As 2q11 is the same band in which the W regions are located, one can speculate that the two regions, the W precursor and *x*de, were transposed to the long arm by the same process. Whereas a copy of the *x*de is still located 23 kb 3′ of C$_\lambda$ (Klobeck and Zachau, 1986), we do not have any evidence that a copy of the W regions exists within the present day *x* locus.

As an alternative to the transposition by a pericentric inversion, the V$_\lambda$ genes could have been transposed via episomes. Such a mechanism has been proposed to be responsible for the amplification of DHFR and mdr1 amplicons in some cell lines (Carroll *et al.*, 1988; Ruiz and Wahl, 1988). While an episome mediated mechanism seems likely for the orphon V$_\lambda$ genes on chromosomes 1, 22 and other chromosomes, we prefer for the W regions the idea of a pericentric inversion.

Data supporting the second assumption, i.e. the new chromosomal location promoted gene amplification, have been reported by Wahl *et al.*, (1984), who found an influence of the chromosomal position of transfected CAD genes on the frequency of CAD amplification.

For the amplification event itself, resulting in the formation of palindromic structures, mechanisms such as those recently reviewed by Stark *et al.* (1989) may be responsible. However, to explain the generation of only one additional copy per amplification step, one has to postulate that no replication processes yielding high copy numbers took place in the course of the W region amplification.

### *Concluding remarks*

Pericentric inversions of chromosome 2 involving the chromosomal segments 2p11−2q13 are observed by cytogenetic methods in about 0.1% of today's population (Djalali *et al.*, 1986 and earlier literature). Also, *de novo* inversions have been observed at chromosome 2 (Vejerslev and Friedrich, 1984), indicating that the inversion is a spontaneous event. We consider the possibility that extensive sequence homologies between parts of the *x* locus on 2p12 and the *x* locus derived W and *x*de regions on 2qcen−q11 promote the inversions. Interstitial telomere-like repeats, found in 2q11−2q14 (Allshire *et al.*, 1988), might also contribute to the spontaneously occurring pericentric inversions observed in the present day human population.

One of the structural features of the W regions, i.e. the organization of different copies as large palindromes, is also found for the *x* locus itself (Lorenz *et al.*, 1987). It is tempting to speculate that a duplication involving large parts of the *x* locus followed the same molecular mechanisms as the amplification of the transposed W precursor. Proving this hypothesis should be feasible by identifying and analysing the junctions between duplicated and non-duplicated sections, and by cloning the head-to-head junction of the two copies forming the *x* locus.

## Materials and methods

### *Recombinant DNA and restriction maps*

The recombinant cosmids were isolated from libraries described by Pohlenz *et al.* (1987) and Lorenz (1989). Colony hybridization was performed as described previously (Klobeck *et al.*, 1987). Restriction maps and subclones were constructed using established methods (Maniatis *et al.*, 1982). Cosmid clones were characterized as described in Pohlenz *et al.* (1987).

### *Nucleic acid hybridization and copy number determination*

DNA transfer was performed according to Reed and Mann (1985), except for PFG, where the protocol of Rigaud *et al.* (1987) was used. Final washing of filters after hybridization was at 68°C with 40 mM phosphate, pH 7.2, 1% sodium dodecylsulphate.

For copy number determinations of amplicons, the insert of clone m654-1/I-1 (Figure 2a) was isolated as a 2.2 kb *Eco*RI−*Bam*HI fragment on an agarose gel and labelled according to Feinberg and Vogelstein (1983). Digested genomic DNA (10 $\mu$g) was electrophoresed, transferred to filters and hybridized. After exposition for 1 day, bands were cut out. For blank values, filter strips above and beneath each band were used. The filter bound activities were measured in a liquid scintillation counter.

### *PFG*

Long-range mapping of DNA from the prostate carcinoma cell line PC-3 (Kaighn *et al.*, 1979) using rare cutting enzymes and orthogonal field gel electrophoresis was performed as described previously (Lorenz *et al.*, 1987).

### *DNA sequencing*

For sequencing of M13-subclones the 'Sequenase' DNA sequencing kit (US Biochemical Corp., Cleveland, OH) was used according to the manufacturer's instructions.

### *Chromosome banding and in situ hybridization*

Metaphase chromosomes were prepared from PHA-stimulated blood lymphocytes. The probes m654-1 and pC-2 were radiolabelled with [$^3$H]dTTP and [$^3$H]dCTP, and hybridization was done as described in Adolph *et al.* (1987). A non-radioactive detection method was used for the localization of cosmid probes. The cosmid clones were linearized and labelled by the random priming technique with biotinylated dUTP-11. Repetitive DNA sequences within the genomic insert were saturated by prehybridization with a 100-fold excess of human Cot-1 DNA for 4 h (Landegent *et al.*, 1987). The normal *in situ* hybridization was performed after prehybridization.

For probe detection the slides must remain humid. They were rinsed in BT buffer (0.1 M sodium bicarbonate, 0.1% Triton-X-100 pH 8.0) and preincubated for 5 min in BT, 0.1% non-fat dry milk. The slides were treated with peroxidase labelled streptavidin (Enzo Biochemicals Inc., New York; 20 $\mu$l per 450 $\mu$l BT) and probe detection was done with diaminobenzidine according to the protocol supplied by the manufacturer. The slides were counter-stained with methylene green.

## Acknowledgements

F.-J.Zimmer was the holder of an A.Butenandt fellowship. We thank Drs G.Bruns and J.Erikson for providing DNA of the cell lines RRP5-3 and J14-2L, respectively, and B.Bauriedel for expert assistance. The work was supported by Bundesministerium für Forschung und Technologie (Center Grant 0316200A2) and Fonds der Chemischen Industrie.

## References

Adolph,S., Bartram,C.R. and Hameister,H. (1987) *Cytogenet. Cell Genet.*, **44**, 65−68.

Allshire,R.C., Gosden,J.R., Cross,S.H., Cranston,G., Rout,D., Sugawara,N., Szostak,J.W., Fantes,P.A. and Hastie,N.D. (1988) *Nature*, **322**, 656−659.

Ardeshir,F., Giulotto,E., Zieg,J., Brison,O., Liao,W.S.L. and Stark,G.R.

(1983) *Mol. Cell. Biol.*, **3**, 2076–2088.
Carroll,S.M., DeRose,M.L., Gaudray,P., Moore,C.M., Needham-Vandevanter,D.R., Von Hoff,D.D. and Wahl,G.M. (1988) *Mol. Cell. Biol.*, **8**, 1525–1533.
Childs,G., Maxson,R., Cohn,R.H. and Kedes,L. (1981) *Cell*, **23**, 651–663.
Djalali,M., Steinbach,P., Bullerdiek,J., Holmes-Siedle,M., Verschraegen-Spae,M.R. and Smith,A. (1986) *Hum. Genet.*, **72**, 32–36.
Erikson,J., Nishikura,K., Ar-Rushdi,A., Finan,J., Emanuel,B., Lenoir,G., Nowell,P.C. and Croce,C.M. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 7581–7585.
Feinberg,A.P. and Vogelstein,B. (1983) *Anal. Biochem.*, **132**, 6–13.
Ford,M. and Fried,M. (1986) *Cell*, **45**, 425–430.
Giulotto,E., Saito,I. and Stark,G.R. (1986) *EMBO J.*, **5**, 2115–2121.
Graninger,W.B., Goldman,P.L., Morton,C.C., O'Brien,S.J. and Korsmeyer,S.J. (1988) *J. Exp. Med.*, **167**, 488–501.
Heartlein and Latt (1989) *Nucleic Acids Res.*, **17**, 1697–1716.
Hyrien,O., Debatisse,M., Buttin,G. and de Saint Vincent,B.R. (1987) *EMBO J.*, **6**, 2401–2408.
Hyrien,O., Debatisse,M., Buttin,G. and de Saint Vincent,B.R. (1988) *EMBO J.*, **7**, 407–417.
Kariya,Y., Kato,K., Hayashizaki,Y., Himena,S., Tarui,S. and Matsubara,K. (1987) *Gene*, **53**, 1–10.
Kaighn,M.E., Narayan,K.S., Ohnuki,Y., Lechner,J.F. and Jones,L.W. (1979) *Invest. Urol.*, **17**, 16–23.
Klobeck,H.-G. and Zachau,H.G. (1986) *Nucleic Acids Res.*, **14**, 4591–4603.
Klobeck,H.-G., Combriato,G. and Zachau,H.G. (1984) *Nucleic Acids Res.*, **12**, 6995–7006.
Klobeck,H.-G., Zimmer,F.-J., Combriato,G. and Zachau,H.G. (1987) *Nucleic Acids Res.*, **15**, 9655–9665.
Landegent,J.E., Jansen in de Wal,N., Dirks,R.W., Baas,F. and van der Ploeg,M. (1987) *Hum. Genet.*, **77**, 366–370.
Lötscher,E., Grzeschik,K.-H., Bauer,H.G., Pohlenz,H.-D., Straubinger,B. and Zachau,H.G. (1986) *Nature*, **320**, 456–458.
Lötscher,E., Zimmer,F.-J., Klopstock,T., Grzeschik,K.-H., Jaenichen,R., Straubinger,B. and Zachau,H.G. (1988) *Gene*, **69**, 215–223.
Lorenz,W. (1989) PhD Thesis, Fakultät für Biologie der Universität München.
Lorenz,W., Straubinger,B. and Zachau,H.G. (1987) *Nucleic Acids Res.*, **15**, 9667–9676.
Ma,L., Looney,J.E., Leu,T.-H. and Hamlin,J.L. (1988) *Mol. Cell. Biol.*, **8**, 2316–2327.
Maeda,N. and Smithies,O. (1986) *Annu. Rev. Genet.*, **20**, 81–108.
Malcolm,S., Barton,P., Murphy,C., Ferguson-Smith,M.A., Bentley,D.L. and Rabbitts,T.H. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 4957–4961.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
McBride,O.W., Hieter,P.A., Hollies,G.F., Swan,D., Otey,M.C. and Leder,P. (1982) *J. Exp. Med.*, **155**, 1480–1490.
Meindl,A. (1990) *PhD Thesis*, Fakultät für Biologie der Universität München.
Nalbantoglu,J. and Meuth,M. (1986) *Nucleic Acids Res.*, **14**, 8361–8371.
Pech,M. and Zachau,H.G. (1984) *Nucleic Acids Res.*, **12**, 9229–9236.
Pohlenz,H.-D. (1986) *PhD Thesis*, Fakultät für Chemie und Pharmazie, Universität München.
Pohlenz,H.-D., Straubinger,B., Thiebe,R., Pech,M., Zimmer,F.-J. and Zachau,H.G. (1987) *J. Mol. Biol.*, **193**, 241–253.
Reed,K.C. and Mann,D.A. (1985) *Nucleic Acids Res.*, **13**, 7207–7221.
Rigaud,G., Grange,T. and Pitet,R. (1987) *Nucleic Acids Res.*, **15**, 857.
Ruiz,J.C. and Wahl,G.M. (1988) *Mol. Cell. Biol.*, **8**, 4302–4313.
Saito,I., Groves,R., Giulotto,E., Rolfe,M. and Stark,G.R. (1989) *Mol. Cell. Biol.*, **9**, 2445–2452.
Shiloh,Y., Shipley,J., Brodeur,G.M., Bruns,G., Korf,B., Donlan,T., Schreck,R.R., Seeger,R., Sakai,K. and Latt,S.A. (1985) *Proc. Natl. Acad. Sci. USA*, **82**, 3761–3765.
Siminovitch,K.A., Bakhshi,A., Goldman,P. and Korsmeyer,S.J. (1985) *Nature*, **316**, 260–262.
Skowronski,J., Fanning,T.G. and Singer,M.E. (1988) *Mol. Cell. Biol.*, **8**, 1385–1397.
Stark,G.R., Debatisse,M., Giulotto,E. and Wahl,G.M. (1989) *Cell*, **57**, 901–908.
Straubinger,B., Thiebe,R., Pech,M. and Zachau,H.G. (1988) *Gene*, **69**, 209–214.
Vejerslev,L.O. and Friedrich,U. (1984) *Prenat. Diagn.*, **4**, 181–186.
Wahl,G.M., de Saint-Vincent,B.R. and Rose,M.L. (1984) *Nature*, **307**, 516–520.
Yeung,C.-Y., Frayne,E.G., Al-Ubaidi,M.R., Hook,A.G., Ingolia,D.E., Wright,D.A. and Kellems,R.E. (1983) *J. Biol. Chem.*, **258**, 15179–15185.
Yunis,J.J. and Prakash,O. (1982) *Science*, **215**, 1525–1530.
Zachau,H.G. (1990) *Biol. Chem. Hoppe-Seyler*, **371**, 1–6.
Zachau,H.G. (1989) In Honjo,T., Alt,F.W. and Rabbitts,T. (eds), *The Immunoglobulin Genes*. Academic Press, London, pp. 91–109.
Zimmer,F.-J. (1989) *PhD Thesis*, Fakultät für Chemie und Pharmazie, Universität München.
Zimmer,F.J., Huber,C., Quenzel,M., Schek,H., Stiller,C., Thiebe,R. and Zachau,H.G. (1990) *Biol. Chem. Hoppe-Seyler*, **371**, 283–293.

*Received on December 27, 1989; revised on February 2, 1990*

J. Mol. Biol. (1995) 247, 536–540

JMB

COMMUNICATION

# SCOP: A Structural Classification of Proteins Database for the Investigation of Sequences and Structures

**Alexey G. Murzin, Steven E. Brenner, Tim Hubbard and Cyrus Chothia***

*MRC Laboratory of Molecular Biology and Cambridge Centre for Protein Engineering, Hills Road Cambridge CB2 2QH England*

*Corresponding author

To facilitate understanding of, and access to, the information available for protein structures, we have constructed the Structural Classification of Proteins (scop) database. This database provides a detailed and comprehensive description of the structural and evolutionary relationships of the proteins of known structure. It also provides for each entry links to co-ordinates, images of the structure, interactive viewers, sequence data and literature references. Two search facilities are available. The homology search permits users to enter a sequence and obtain a list of any structures to which it has significant levels of sequence similarity. The key word search finds, for a word entered by the user, matches from both the text of the scop database and the headers of Brookhaven Protein Databank structure files. The database is freely accessible on World Wide Web (WWW) with an entry point to URL http://scop.mrc-lmb.cam.ac.uk/scop/

*scop: an old English poet or minstrel (Oxford English Dictionary);*

*cxon: pile, accumulation (Russian Dictionary).*

*Keywords:* protein families; superfamilies; folds; evolutionary relationships

Nearly all proteins have structural similarities with other proteins and, in many cases, share a common evolutionary origin. The knowledge of these relationships makes important contributions to molecular biology and to other related areas of science. It is central to our understanding of the structure and evolution of proteins. It will play an important role in the interpretation of the sequences produced by the genome projects and, therefore, in understanding the evolution of development.

The recent exponential growth in the number of proteins whose structures have been determined by X-ray crystallography and NMR spectroscopy means that there is now a large and rapidly growing corpus of information available. At present (January, 1995) the Brookhaven Protein Databank (PDB, (Abola *et al.*, 1987)) contains 3091 entries and the number is increasing by about 100 a month. To facilitate the understanding of, and access to, this information, we have constructed the Structural Classification of Proteins (scop) database. This database provides a detailed and comprehensive description of the structural and evolutionary relationships of proteins whose three-dimensional structures have been determined. It includes all proteins in the current version of the PDB and almost all proteins for which structures have been published but whose co-ordinates are not available from the PDB.

The classification of protein structures in the database is based on evolutionary relationships and on the principles that govern their three-dimensional structure. Early work on protein structures showed that there are striking regularities in the ways in which secondary structures are assembled (Levitt & Chothia, 1976; Chothia *et al.*, 1977) and in the topologies of the polypeptide chains (Richardson, 1976, 1977; Sternberg & Thornton, 1976). These regularities arise from the intrinsic physical and chemical properties of proteins (Chothia, 1984; Finkelstein & Ptitsyn, 1987) and provide the basis for the classification of protein folds (Levitt & Chothia, 1976; Richardson, 1981). This early work has been taken further in more recent papers; see, for example, Holm & Sander (1993), Orengo *et al.* (1993), Overington *et al.* (1993) and Yee & Dill (1993). An extensive bibliography of papers on the classification and the determinants of protein folds is given in scop.

The method used to construct the protein classification in scop is essentially the visual inspection and comparison of structures though various automatic tools are used to make the task manageable and help provide generality. Given the

Abbreviations used: PDB, Protein Databank; scop, Structural Classification of Proteins.

0022-2836/95/140536-05 $08.00/0

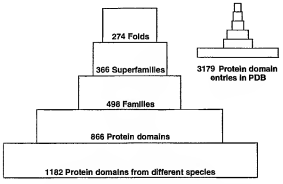



**Figure 1.** In scop, the unit of classification is usually the protein domain. Small proteins, and most of those of medium size, have a single domain and are, therefore, treated as a whole. The domains in large proteins are usually classified individually. The protein entries in the December 1994 of the Brookhaven Protein Databank (PDB) contain 3179 domains. Many of these become forms of the same protein whose differences are not significant in terms of the classification used here; for example they have different bound ligands or engineered mutations. To distinguish between these and structures of the same protein from different organisms, proteins listed within a family are subclassified by species. Classification of the 3179 domains show that they come from 498 families that can be clustered into 366 superfamilies and 279 different folds. In addition to these, scop contains entries for 195 proteins that do not have atomic co-ordinates available from the PDB at present but for which description of their structures have been published.

current limitations of purely automatic procedures, we believe this approach produces the most accurate and useful results. The unit of classification is usually the protein domain. Small proteins, and most of those of medium size, have a single domain and are, therefore, treated as a whole. The domains in large proteins are usually classified individually.

The classification is on hierarchical levels that embody the evolutionary and structural relationships.

**FAMILY.** Proteins are clustered together into families on the basis of one of two criteria that imply their having a common evolutionary origin: first, all proteins that have residue identities of 30% and greater; second, proteins with lower sequence identities but whose functions and structures are very similar; for example, globins with sequence identities of 15%.

**SUPERFAMILY.** Families, whose proteins have low sequence identities but whose structures and, in many cases, functional features suggest that a common evolutionary origin is probable, are placed together in superfamilies; for example, actin, the ATPase domain of the heat-shock protein and hexokinase (Flaherty *et al.*, 1991).

**COMMON FOLD.** Superfamilies and families are defined as having a common fold if their proteins have same major secondary structures in same arrangement with the same topological connections. In scop we give for each fold short descriptions of its main structural features. Different proteins with the same fold usually have peripheral elements of secondary structure and turn regions that differ in size and conformation and, in the more divergent cases, these differing regions may form half or more of each structure. For proteins placed together in the same fold category, the structural similarities probably arise from the physics and chemistry of proteins favouring certain packing arrangements and chain topologies (see above). There may, however, be cases where a common evolutionary origin is obscured by the extent of the divergence in sequence, structure and function. In these cases, it is possible that the discovery of new structures, with folds between those of the previously known structures, will make clear their common evolutionary relationship.

**CLASS.** For convenience of users, the different folds have been grouped into classes. Most of the folds are assigned to one of the five structural classes on the basis of the secondary structures of which they composed: (1) all alpha (for proteins whose structure is essentially formed by α-helices), (2) all beta (for those whose structure is essentially formed by β-sheets), (3) alpha and beta (for proteins with α-helices and β-strands that are largely interspersed), (4) alpha plus beta (for those in which α-helices and β-strands are largely segregated) and (5) multi-domain (for those with domains of different fold and for which no homologues are known at present). Note that we do not use Greek characters in scop because they are not accessible to all world wide web viewers. More unusual proteins, peptides and the PDB entries for designed proteins,

**Table 1**

Facilities and databases to which SCOP has links

| Link | Source | URL | Reference |
|---|---|---|---|
| Co-ordinates | PDB | http://www.pdb.bnl.gov/ | (Abola *et al.*, 1987) |
| Static images | SP3D | http://expasy.hcuge.ch/ gopher://pdb.pdb.bnl.gov/ | (Appel *et al.*, 1994) |
| On-the-fly images | NIH molecular modelling group | http://www.nih.gov/www94/molrus | (FitzGerald, 1994) |
| Sequences and MEDLINE entries | NCBI Entrez | http://www.ncbi.nlm.nih.gov/ | (Benson *et al.*, 1993) |

The scop database contains links to a number of other facilities and databases in the world. Several interactive viewers can be linked with scop using PDB co-ordinates. The location and nature of the links will vary as databases evolve and relocate.

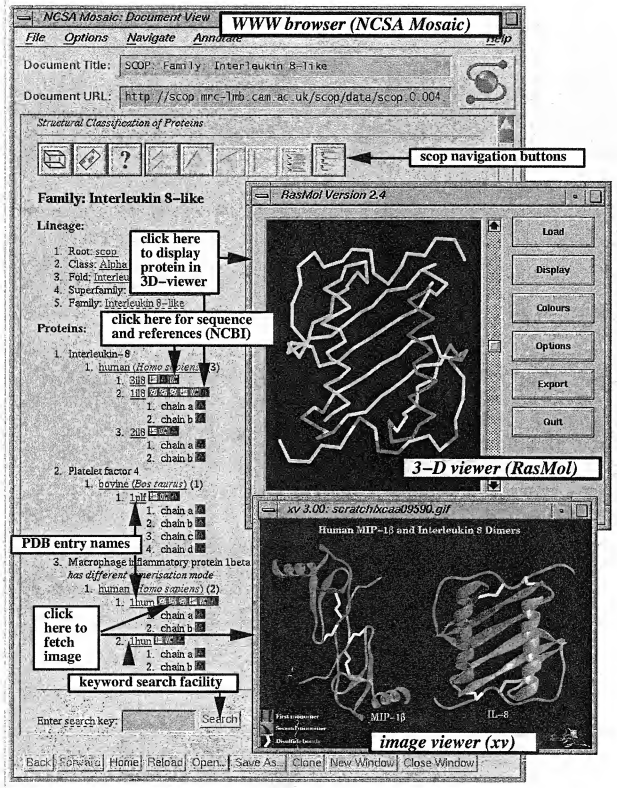



Figure 2. A typical scop session is shown on a unix workstation. A scop page, of the Interleukin 8-like family, is displayed by the *WWW browser program (NCSA Mosaic)* (Schatz & Hardin, 1994). Navigating through the tree structure is accomplished by selecting any underlined entry, by clicking on buttons (at the top of each page) and by keyword searching (at the bottom of each page). The static image comparing two proteins in this family was downloaded by clicking on the icon indicated and is displayed by image-viewer program *xv*. By clicking on one of the green icons, commands were sent to a molecular viewer program (*RasMol*) written by Roger Sayle (Sayle, 1994), instructing it to automatically display the relevant PDB file and colour the domain in question by secondary structure. Since sending large PDB files over the network can be slow, this feature of scop can be configured to use local copies of PDB files if they are available. Equivalent WWW browsers, image-display programs and molecular viewers are also available free for Windows-PC and Macintosh platforms.

theoretical models, nucleic acids and carbohydrates, have been assigned to other classes.

The number of entries, families, superfamilies and common folds in the current version of scop are shown in Figure 1. The exact position of boundaries between family, superfamily and fold are, to some degree, subjective. However, because all proteins that could conceivably belong to a family or superfamily are clustered together in the encompassing fold category, some users may wish to concentrate on this part of the database.

In addition to the information on structural and evolutionary relationships, each entry (for which co-ordinates are available) has links to images of the structure, interactive molecular viewers, the atomic co-ordinates, sequence data and homologues and MEDLINE abstracts (see Table 1).

Two search facilities are available in scop. The homology search permits users to enter a sequence and obtain a list of any structures to which it has significant levels of sequence similarity. The key word search finds, for a word entered by the user, matches from both the text of the scop database and the headers of Brookhaven Protein Databank structure files.

To provide easy and broad access, we have made the scop database available as a set of tightly coupled hypertext pages on the world wide web (WWW). This allows it to be accessed by any machine on the internet (including Macintoshes, PCs and workstations) using free WWW reader programs, such as Mosaic (Schatz & Hardin, 1994). Once such a program has been started, it is necessary only to "open" URL:

http://scop.mrc-lmb.cam.ac.uk/scop/

to obtain the "home" page level of the database.

In Figure 2 we show a typical page from the database. Each page has buttons to go back to the top-level home page, to send electronic mail to the authors, and to retrieve a detailed help page. Navigating through the tree structure is simple; selecting any entry retrieves the appropriate page. In addition, buttons make it possible to move within the hierarchy in other manners, such as "upwards" to obtain broader levels of classification.

The scop database was originally created as a tool for understanding protein evolution through sequence-structure relationships and determining if new sequences and new structures are related to previously known protein structures. On a more general level, the highest levels of classification provide an overview of the diversity of protein structures now known and would be appropriate both for researchers and students. The specific lower levels should be helpful for comparing individual structures with their evolutionary and structurally related counterparts. In addition, we have also found that the search capabilities with easy access to data and images make scop a powerful general-purpose interface to the PDB.

As new structures are released by PDB and published, they will be entered in scop and revised versions of the database will be made available on WWW. Moreover, as our formal understanding of relationships between structure, sequence function and evolution grows, it will be embodied in additional facilities in the database.

*Acknowledgements*

We thank Sean Eddy, Graeme Mitchison and Erik Sonnhammer for discussions and useful suggestions and Roger Sayle, the author of rasmol, for suggesting the tcl/tk interface to rasmol. The University of Cambridge School of Biological Sciences is providing the principal database access point. S.E.B. is grateful to Herchel Smith and Harvard University, St. John's College, Cambridge Overseas Trust, American Friends of Cambridge University and CVCP/ORS for support. T.H. is grateful to ZENECA for support.

## References

Abola, E., Bernstein, F. C., Bryant, S. H., Koetzle, T. F. & Weng, J. (1987). Protein Data Bank. In *Crystallographic Databases—Information Content, Software Systems, Scientific Applications* (Allen, F. H., Bergerhoff, G. & Sievers, R., eds), pp. 107–132, Commission of the International Union of Crystallography, Bonn, Cambridge, Chester.

Appel, R. D., Bairoch, A. & Hochstrasser, D. F. (1994). A new generation of information retrieval tools for biologists: the example of the ExPASy WWW server. *Trends Biochem. Sci.* 19, 258–260.

Benson, D., Lipman, D. J. & Ostell, J. (1993). Genbank. *Nucl. Acids Res.* 21, 2963–2965.

Chothia, C. (1984). Principles that determine the structure of proteins. *Annu. Rev. Biochem.* 53, 537–572.

Chothia, C., Levitt, M. & Richardson, D. (1977). Structure of proteins: packing of α-helices and β-sheets. *Proc. Nat. Acad. Sci., U.S.A.* 74, 4130–4134.

Finkelstein, A. V. & Ptitsyn, O. B. (1987). Why do globular proteins fit the limited set of folding patterns. *Prog. Biophys. Mol. Biol.* 50, 171–190.

FitzGerald, P. C. (1994). A WWW Forms interface to facilitate access (browsing, searching and viewing) of the molecular structure data contained within the Brookhaven Protein Data Bank (PDB). *Proceedings of WWW94 (First International Conference on the World Wide Web), Chemistry Workshop, CERN, Geneva,* Elsevier Science BV, Switzerland.

Flaherty K. M., McKay, D. B., Kabsch, W. & Holmes, K. C. (1991). Similarity of the three-dimensional structures of actin and the ATPase fragment of a 70 kDa heat shock cognate protein. *Proc. Nat. Acad. Sci., U.S.A.* 88, 5041–5045.

Holm, L. & Sander, C. (1993). Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* 233, 123–138.

Levitt, M. & Chothia, C. (1976). Structural patterns in globular proteins. *Nature (London)*, 261, 552–558.

Orengo, C., Flores, T. P., Taylor, W. R. & Thornton, J. M. (1993). Identifying and classifying protein fold families. *Protein Eng.* 6, 485–500.

Overington, J. P., Zhu, Z. Y., Sali, A., Johnson, M. S., Sowdhamini, R., Louie, C. & Blundell, T. L. (1993). Molecular recognition in protein families: a database of three-dimensional structures of related proteins. *Biochem. Soc. Trans.* 21, 597–604.

Richardson, J. S. (1976). Handedness of crossover connections in β-sheets. *Proc. Nat. Acad. Sci., U.S.A.* **73**, 2619–2623.

Richardson, J. S. (1977). β-Sheet topology and the relatedness of proteins. *Nature (London)*, **268**, 495–500.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.

Sayle, R. (1994). Rasmol. *WWW, URL ftp://ftp.dcs.ed.ac.uk/rasmol.*

Schatz, B. R. & Hardin, J. B. (1994). NCSA Mosaic and the world wide web: global hypermedia protocols for the Internet. *Science*, **265**, 895–901.

Sternberg, M. J. E. & Thornton, J. M. (1976). On the conformation of proteins: the handedness of the β-strand–α-helix–β-strand unit. *J. Mol. Biol.* **105**, 367–382.

Yee, D. P. & Dill, K. A. (1993). Families and the structural relatedness among globular proteins. *Protein Sci.* **2**, 884–899.

*Edited by F. E. Cohen*

*(Received 1 November 1994; accepted 11 January 1995)*

*Proc. Nat. Acad. Sci. USA*
Vol. 72, No. 7, pp. 2785–2788, July 1975
Immunology

# Evolutionary and structural influences on light chain constant ($C_L$) region of human and mouse immunoglobulins

**(structure-function correlation and evolution/λ–κ comparisons/comparative structures)**

ELVIN A. KABAT*, EDUARDO A. PADLAN†, AND DAVID R. DAVIES†

* Fogarty International Center and † Laboratory of Molecular Biology, National Institute of Arthritis, Metabolism and Digestive Diseases, National Institutes of Health, Bethesda, Maryland 20014; and * Departments of Microbiology, Human Genetics and Development and Neurology, Columbia University and the Neurological Institute, Presbyterian Hospital, New York, N.Y. 10032

*Contributed by Elvin A. Kabat, April 17, 1975*

ABSTRACT A comparison of five constant region sequences of human and mouse κ and λ immunoglobulin chains has been undertaken in order to reveal sequence homologies and evolutionary relationships. Simultaneously, a comparison with the three-dimensional structure of one mouse κ-chain (McPC 603) has suggested structural reasons why many of the residues are invariant or conserved along κ versus λ lines. There are a number of residues that have remained invariant despite exposed positions for reasons that do not appear to be connected with the folding of this $C_L$ domain.

The constant region ($C_L$ domain) of immunoglobulin light (L) chains contributes substantially to the functioning of an immunoglobulin molecule. While it is not involved directly in the specificity and complementarity of the antibody combining sites, it is joined to its counterpart in the heavy (H) chain, the $C_H1$ domain, by a variety of noncovalent interactions as well as in most immunoglobulins by a disulfide bond at its C-terminal or subterminal Cys. This –S–S– bond is not essential to L–H association, which is maintained noncovalently even after reduction and alkylation. In one immunoglobulin subclass of IgA2, the L–H bond does not occur, but an L–L dimer is formed, which remains noncovalently linked to the H chains. Bence Jones proteins in the form of L–L dimers also occur and may be held together by noncovalent forces (1).

There are two subclasses of light chains, κ and λ, which are present in almost all species examined; both are found in all five classes of immunoglobulins, IgG, IgM, IgA, IgD, and IgE, but each immunoglobulin molecule contains two identical κ or two identical λ chains.

The availability of the sequences of the $C_L$ domain of human κ, human λ, mouse κ, and two mouse λ light chains (2), together with x-ray data on the three-dimensional structure of this domain (3–5) made it desirable to evaluate, if possible, structural influences on the evolution of these domains.

The present study attempts, residue by residue among these five chains, to relate preservation or variation of sequence to structure and function as evaluated from a three-dimensional model of the $C_L$ and its interactions with the $C_H1$ domain. The findings show some interesting stretches of sequence in which invariance predominates, and others in which evolutionary divergence has been essentially along κ and λ lines. Positions at the surface of protein molecules that are accessible to solvent may undergo many mutational changes which do not affect three-dimensional folding, while residues that are buried tend to be invariant or highly conserved (6). In $C_L$ domains mutations involving residues that contact the $C_H1$ domain also tend to be restricted.

The present study shows that residues preserved along κ and λ lines generally show conservative substitutions if in the interior of the domain or if buried. Fewer exposed residues which have diverged along κ and λ lines are homologous. Certain residues remain invariant despite an essentially exposed position and for no obvious reason. Of special interest is the observation that at only two and four positions, respectively, were human κ identical with human λ and mouse κ identical with mouse λ, while human κ and mouse κ were identical at 29 and human λ and mouse λ at 39 positions.

## MATERIALS AND METHODS

The model of the Fab fragment of mouse McPC 603 constructed from x-ray data at 3.1 Å resolution was used (5) as well as published information on the $C_L$ regions of a Bence Jones dimer (4) and human Fab fragment (3). Sequences of human κ, human λ, mouse κ, and two mouse λ chains were available (2). These sequences were aligned for maximum structural and sequence homology from residues 101 to 215 and modified to include the additional data reported (7). Each residue was located in the model of McPC 603 and classified according to whether it was: exposed, 0; mainly exposed, 1; partly exposed, partly buried, 2; mainly buried, 3; completely buried, 4; or in contact with $C_H1$, C. In addition, each position was classified as: invariant; four of five chains and three of five chains identical; human κ and human λ identical; mouse κ and mouse λ identical; human and mouse κ identical; human and mouse λ identical; and human κ, human λ, mouse κ and mouse λ different.

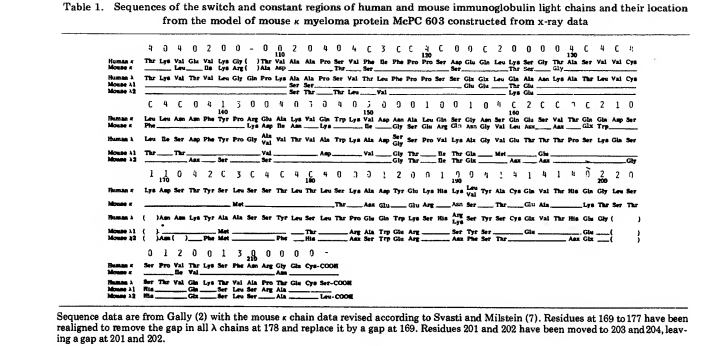
## RESULTS

Table 1 lists the sequences of the five chains from positions 101 to 215. Above each residue is its classification from its position in the model.

Fig. 1 summarizes the sequence data in Table 1 with respect to the identities specified above. It is evident that clusters of invariant residues and those with 4/5 chains identical occur, notably at positions 118 to 123, 148 to 152 (excluding 150), 176 to 182 (180 3/5 identical), and 194 to 200 (exclud-

Abbreviations: C and V, constant and variable regions of immunoglobulin chains; L and H, light and heavy chains of immunoglobulins.

Table 1. Sequences of the switch and constant regions of human and mouse immunoglobulin light chains and their location from the model of mouse $\kappa$ myeloma protein MePC 603 constructed from x-ray data



Sequence data are from Gally (2) with the mouse $\kappa$ chain data revised according to Svasti and Milstein (7). Residues at 169 to 177 have been realigned to remove the gap in all $\lambda$ chains at 178 and replace it by a gap at 169. Residues 201 and 202 have been moved to 203 and 204, leaving a gap at 201 and 202.

ing 195). There are also several clusters in which $\kappa$ versus $\lambda$ diversification predominates, 128 to 132 (130 invariant) and especially 164 to 172.

Of the 114 residues considered, 28 were invariant, 17 and 8, respectively, showed 4/5 and 3/5 chains identical. Of the kinds of identities of two chains in the remaining positions human $\kappa$ and human $\lambda$ were identical at only two positions and mouse $\kappa$ and mouse $\lambda$ at only four positions, while human and mouse $\kappa$ were the same at 29 positions and human $\lambda$ and one or both mouse $\lambda$ chains at 39 positions. At two positions a human $\kappa$ and a mouse $\lambda$ chain were the same (*) and at three positions human $\lambda$ and mouse $\kappa$ chains had the same residue (●). At but eight positions were all four chains different.

Table 2 summarizes the sequence data in Fig. 1 in relation to the location of each residue in the model. The most striking finding is that of the eight positions at which the four chains had different amino acids (Fig. 1), six residues were completely exposed to solvent and the remaining two were mainly exposed. Moreover, of the two positions at which human $\kappa$ and human $\lambda$ were the same (one of which was the Oz marker at position 190) and of the four positions at which mouse $\kappa$ and mouse $\lambda$ were identical, all but one were completely or mainly exposed. The sixth, residue 135, was a contacting residue adjacent to invariant Cys 134. Three of the five positions in which human and mouse identities occurred but which were not both $\kappa$ or both $\lambda$ as indicated by the symbols (* and ●) in Fig. 1 were completely exposed;

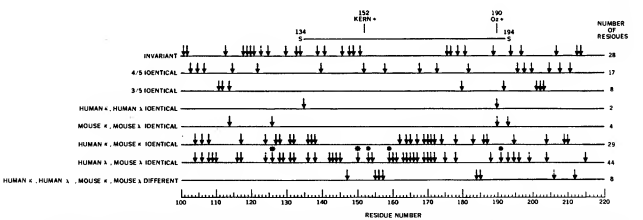



FIG. 1. Distribution of identical residues in the switch and constant regions of human and mouse immunoglobulin light chains. The arrows indicate identical residues in two or more of the five chains as well as at positions at which the four classes of chains differed. When more than one arrow occurs at a given position, there was identity among two sets of chains. Thus, at position 169 human and mouse $\kappa$ chains had Lys while human $\lambda$ and both mouse $\lambda$ chains had a gap. Arrows with an * and a ● indicate the few residues identical in human $\kappa$ and mouse $\lambda$ and human $\lambda$ and mouse $\kappa$, respectively. A dashed arrow indicates an Asx or Glx at that position in one or more chains.

Table 2. Location from x-ray structure and degree of evolutionary preservation of residues in the $C_L'$ domain and switch. region of human and mouse light chains

| | | Identities—number of residues | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Location | Code | Invariant | 4/5 | 3/5 | Human κ Human λ | Mouse κ Mouse λ | Human κ Mouse κ | Human λ Mouse λ | Human κ, Human λ Mouse κ, Mouse λ different |
| Exposed | 0 | 7 | 7 | 3 | 1 | 2 | 10 | 16 | 6 |
| Mainly exposed | 1 | 2 | 2 | | | 1 | 4 | 4 | 2 |
| Partly exposed, partly buried | 2 | 1 | 4 | 1 | | | 2 | 3 | |
| Mainly buried | 3 | 2 | | 2 | | | 2 | 1 | |
| Completely buried | 4 | 11 | 4 | 1 | | | 3 | 5 | |
| Contacts H chain | C | 5 | | 1 | 1 | 1 | 8 | 7 | |
| Total | | 28 | 17 | 8 | 2 | 4 | 29 | 36‡ | 8 |

‡ Excludes residues with * and • in Fig. 1 and residue 215 not specifiable; gaps in κ at position 108 and in λ chains at positions 169 and 201 and 202.

these included the Inv marker 191. Position 152, which in human λ chains carries the Kern marker, was also exposed (3).

Of the 28 invariant residues 18 were contact residues or were mainly or completely buried, while nine were mainly or completely exposed. Considering together the 25 positions in which 4/5 and 3/5 residues were identical, 12 were completely or mainly exposed, only one was a contact residue, while five were completely buried.

Of the 29 positions at which human κ and mouse κ had the same amino acid and the six (footnote Table 2) at which human λ and one of the mouse λ chains had the same amino acid, seven in each were contacting residues and three and six, respectively, were completely buried while 15 and 20, respectively, were completely or mainly exposed.

Of the seven invariant exposed residues six were charged, 2 Glu, 1 Asp, and 3 Lys; the seventh was Cys 214 which forms the –S–S– bond to the H-chain or in some cases to another L-chain. The location of these six residues and the two mainly exposed residues His and Thr in the model provides no insight into the basis for their invariance.

The remaining 19 invariant residues, those partly or completely buried and the contacting residues, included 12 strongly hydrophobic residues, 1 Trp, 2 Phe, 4 Pro, 3 Leu, and 2 Val; the remaining residues were Cys 134 and Cys 194, which form the domain S–S bond, two less strongly hydrophobic residues Ala and Thr, and three Ser.

Examining the residues with 4/5 and 3/5 identities, the partly, completely buried, and contacting residues were also overwhelmingly hydrophobic, consisting of 2 Tyr, 5 Val, 1 Thr, and 1 Ala, the others being 1 Gly and 1 His and the gap at position 201. The nonidentical residues at these positions were also hydrophobic in almost all cases, involving replacements of 2 Ile, 2 Leu, and Ala for the five Val residues and of Phe for one Tyr; the remaining Tyr 192 was replaced by Ser in both mouse λ chains, the Thr by Ser or His, the Ala by Asx, and the Gly by Thr. There is obviously an extraordinary preservation of structure in terms of these groups of residues.

Among the 15 and 16 residues that have diverged along κ versus λ lines and which are partly, mainly, and completely buried or are contacting residues, eight are identical pairs; three of these involve conservative hydrophobic substitutions, Ile-Leu, Val-Leu, and Leu-Ile, at positions 117, 132, and 136, respectively, involving mainly and completely buried residues; the remaining pairs are the substitutions Thr-Lys at 172, Gln-Glu 124, Ser-Thr 131, Glu-Ser 165, and Thr-Tyr 178, the last four being contacting residues. The presence in κ chains of an additional residue at 169 causes Thr 172 to be completely buried in the mouse $C_\kappa$ domain; in the λ chains, Lys 172 is probably completely exposed to solvent. The unpaired residues are 1 Phe, 1 Tyr, 2 Ser, 1 Asp, and 1 Thr in the κ, and 1 Ala, 1 Val, 1 Gln, 1 Glu, 1 Lys, and 3 Thr in the λ group.

The κ versus λ differences show a predominance of completely and mainly exposed residues with 14 of 29 and 20 of 36 residues in κ and λ, respectively, falling into this group; nine are pairs.

The region 101–108 consists of two invariant residues, 101 and 102, followed by six positions in which residues with 4/5 identities alternate with residues which have evolved along κ and λ lines, including the gap at position 108; four of the positions are completely exposed, one is partly exposed, and two are completely buried. Arg 107 marks the end of the mouse $V_\kappa$ domain; $C_\kappa$ starts with Ala 109. The additional residue at position 108 in λ chains could be accommodated by a hairpin bend facilitated by Gly 107 or by Pro 109.

There is a cluster of invariant residues from 118 to 123 consisting of three contacting residues, 118, 119, and 121, one partly buried 120, and two exposed residues, 121 Ser (4/5) with an Asp alternative, and an invariant Glu 123. The region 124–140 contains some invariants but is largely made

up of residues which have evolved along κ and λ lines; residues 130–137 consist exclusively of contacting and completely buried residues of which three are invariant, in another three κ and λ differ, and in the others more variation has occurred. At 135 human κ and λ have Leu while mouse κ has Phe and mouse λ Thr and at position 137 human and mouse κ have Asn, human λ Ser, and mouse λ Thr.

The most striking region which has been preserved along κ versus λ lines, 160 to 175, has five contacting, one completely and four partly buried residues, and two mainly and three completely exposed residues. It is followed by a largely invariant cluster, 176 to 181, of contacting and buried residues. The remainder of the molecule is mainly exposed except for the region around Cys 194, in which buried and mainly exposed residues alternate and there is no clustering of invariant or κ versus λ residues.

## DISCUSSION

The general structural relationships described for the $C_L$ region of human and mouse κ and λ chains are an initial attempt to understand the basis for the evolutionary preservation of certain regions as essentially invariant and the preservation of others along κ versus λ lines, an evolutionary divergence which took place about 200 million years ago (8). The principles established from sequence and structural findings on other proteins generally apply equally well to the immunoglobulins. Thus the buried residues and those contacting the heavy chain tend to be largely invariant and hydrophobic, while residues that are exposed or mainly exposed may vary, and are generally polar. These also include the few residues that have evolved along human versus mouse lines. There are, however, substantial numbers of hydrophobic residues that are invariant or identical in 4/5 or 3/5 chains and that occur in regions of the molecule for which no obvious structural basis for their conservation may be assigned.

The marked clustering of residues which have been preserved along κ versus λ lines in certain portions of the chain, most notably at positions 160 to 175, or have been maintained invariant, such as 118 to 123, suggests that these may have unique functions.

The $C_L$ domain has been extraordinarily preserved once κ versus λ diversification occurred, since at only eight positions were all four chains different and at only five other positions were human and mouse chains identical despite a κ versus λ difference (Fig. 1).

The immunoglobulin findings were compared with values for human and mouse hemoglobins, using the α chain as equivalent to κ and the β (γ) chain equivalent to λ (δ). Since hemoglobin chains are longer, the data were normalized to 115 residues. The findings for hemoglobin are strikingly different from the immunoglobulin results. Thirty-six residues were invariant, and 4/5 and 3/5 chains were identical at 12 and at 49 positions, respectively. Human and mouse α and human and mouse β were identical at 13 and 11 positions, respectively. At only one position each were mouse α and mouse β, human α and mouse β, and human β and mouse α identical. There were no positions at which all four chains differ. Thus there were 97 residues in which three or more chains were identical in hemoglobins in contrast to 53 such residues in the immunoglobulins. This comparison indicates a higher degree of evolutionary conservation of residues that did not differentiate along α versus β lines than residues that differentiated along κ versus λ lines. The immunoglobulin $C_L$ domain thus shows a higher degree of evolutionary adaptability than do the hemoglobins.

1. Kabat, E. A. (1968) *Structural Concepts in Immunology and Immunochemistry* (Holt, Rinehart and Winston, New York), 305 pp.
2. Gally, J. A. (1973) in *The Antigens*, ed. Sela, M. (Academic Press, New York), pp. 161–298.
3. Poljak, R. J., Amzel, L. M., Chen, B. L., Phizackerley, R. P. & Saul, F. (1974) *Proc. Nat. Acad. Sci. USA* 71, 3440–3444.
4. Edmundson, A. B., Ely, K. R., Girling, R. L., Abola, E. E., Schiffer, M. & Westholm, F. A. (1974) in *Progress in Immunology II, Vol. 1: Immunochemical Aspects*, eds. Brent, L. & Holborow, J. (North-Holland Publishing Co., Amsterdam), pp. 103–113.
5. Segal, D. M., Padlan, E. A., Cohen, G. H., Rudikoff, S., Potter, M. & Davies, D. R. (1974) *Proc. Nat. Acad. Sci. USA* 71, 4298–4302.
6. Perutz, M. F., Kendrew, J. C. & Watson, H. C. (1965) *J. Mol. Biol.* 13, 669–678.
7. Svasti, J. & Milstein, C. (1972) *Biochem. J.* 128, 427–444.
8. Dayhoff, M. O., ed. (1972) *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C.).

doi:10.1016/j.jmb.2006.10.093 J. Mol. Biol. (2007) 367, 603–608

JMB

Available online at www.sciencedirect.com
ScienceDirect

ELSEVIER

COMMUNICATION

# β-Edge Interactions in a Pentadecameric Human Antibody Vκ Domain

**Leo C. James[1]*, Phil C. Jones[2], Airlie McCoy[3], Glenys A. Tennent[4] Mark B. Pepys[4], Kristoffer Famm[5] and Greg Winter[1]***

[1]*MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK*

[2]*Domantis Limited, 315 Cambridge Science Park Cambridge, CB4 0WG, UK*

[3]*Cambridge Institute for Medical Research, Wellcome Trust/MRC Building Addenbrooke's Hospital, Hills Road, Cambridge, CB2 0XY UK*

[4]*Centre for Amyloidosis and Acute Phase Proteins Department of Medicine Royal Free and University College Medical School London NW3 2PF, UK*

[5]*MRC Centre for Protein Engineering, Hills Road Cambridge CB2 2QH, UK*

**Corresponding authors*

Antibodies are the archetypal molecules of the Ig-fold superfamily. Their highly conserved β-sheet architecture has evolved to avoid aggregation by protecting edge strands. However, the crystal structure of a human Vκ domain described here, reveals an exposed β-edge strand which mediates assembly of a helical pentadecameric oligomer. This edge strand is highly conserved in Vκ domains but is both shortened and capped by the use of two sequential *trans*-proline residues in Vλ domains. We suggest that the exposure of this β-edge in Vκ domains may explain why light-chain deposition disease is mediated predominantly by κ antibodies.



*Keywords:* β-edge; amyloid; aggregation; antibody; light-chain deposition disease

Antibody immunoglobulin variable domains are amongst the most intensively studied of β-protein structures. They are comprised of two chains, light and heavy, which form a functional heterodimer. Each chain in turn is comprised of several domains, the N-terminal variable (V) domain of which is responsible for targeting to pathogens and is the focus of somatic hypermutation. Human antibodies utilise two alternative light-chain sequences, subtypes κ and λ, which appear to be functionally identical and are highly structurally homologous. Nevertheless, there are pathological differences. λ antibodies are found in two-thirds of light-chain amyloidosis (AL) cases[1] whereas κ antibodies mediate >85% of light-chain deposition disease.[2,3] Furthermore, κ antibodies can be targeted by the bacterial pathogen *Peptostreptococcus magnus* superantigen, Protein L, whereas λ cannot.[4]

We wondered whether these differences were due to structural differences between the isotypes. Superposition of the X-ray crystallographic structures of the κ and λ domains reveals that they are highly similar and broadly duplicate the same topological features. Each variable domain has a typical immunoglobulin fold consisting of two antiparallel β-sheets closely packed to produce a flattened β-barrel (Figures 1 and 2). As in other β-sheet containing proteins, antibodies employ "negative design" features to protect edge strands and prevent aggregation of the native state.[5] In antibody light-chains, one end of the β-barrel is formed by variable loop CDR2, which twists in an S shape to

Abbreviation used: CDR, complementarity-determing region.

E-mail addresses of the corresponding authors: lcj@mrc-lmb.cam.ac.uk; winter@mrc-lmb.cam.ac.uk

0022-2836/$ - see front matter © 2007 Published by Elsevier Ltd.

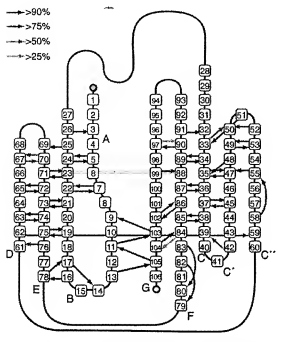



**Figure 1.** 2-D view of the Vκ domain secondary structure. The domain has been divided along the 2-fold and opened out with the "outer" sheet on the left and "inner" sheet on the right. Main-chain hydrogen bonds are marked according to their conservation amongst Vκ sequences.

cap both strand C′ on the five-strand inner sheet and strand D on the three-strand outer sheet. At the other end of the barrel, N-terminal strand A employs a "strand switch" device to protect the two edge strands from each sheet (B on the outer-sheet and G on the inner; Figure 1).

It is in the design of this strand switch that κ and λ clearly diverge. Vκ domains force a strand switch through an absolutely conserved *cis*-proline at position 8 (Figures 1 and 2), leaving the post-kink strand structurally exposed at residues 9–12. Indeed this region is able to mediate intermolecular β-zipper interactions with the *P. magnus* protein, Protein L (PpL). By contrast Vλ chains combine a *trans*-proline at residue 8 with an additional *trans*-proline at position 9, whereby the post-kink strand is shortened to three residues and capped through the planar ring of proline 9.

During studies on Vκ domains, gel filtration of VD9[6] revealed multiple oligomeric species, and it spontaneously aggregated at concentrations above 4 mg/ml. The aggregates were visibly amorphous but when examined under the electron microscope resolved as straight non-branching fibres 7.1(±0.3) nm in width and 60–180 nm long (Supplementary Data, Figure 1). Whilst these fibres superficially resemble amyloid, Congo red and serum amyloid P component binding assays revealed that amyloid was not present in appreciable amounts (Supplementary Data, Figure 2 and Supplementary Data, Table 1). When concentrated to between 3–4 mg/ml and allowed to concentrate slowly by vapour diffusion against a wide range of different buffers VD9 precipitated in regular hexameric crystals. We

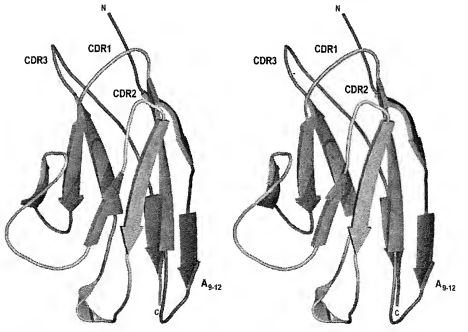



**Figure 2.** Stereo view of a Vκ domain. Secondary structure representation of Vκ VD9 showing CDR loops and β-edge strand A. View is rotated 90° with respect to Figure 1.

**Table 1.** Data collection and structure refinement statistics

| | VD9 |
|---|---|
| Space group | $P6_122$ |
| Cell (Å) | $a=b=191.9$, $c=197.3$ |
| Number per a.u. | 15 |
| Resolution (Å) | 2.6 |
| Unique reflections | 62,507 |
| $R_{merge}$ | 0.093 (0.421) |
| Redundancy | 2.3 (1.9) |
| Completeness (%) | 91.4 (89.7) |
| Average I/σI | 8.5 (1.9) |
| Final $R_{cryst}$ | 0.25 |
| Final $R_{free}$ | 0.29 |
| Bond r.m.s.d. | 0.009 |
| Angle r.m.s.d. (deg.) | 1.5 |
| Ramachandran statistics (%) | |
| Favoured | 85.8 |
| Allowed | 14.2 |

Numbers in parentheses are for the highest resolution shell.

used these crystals to determine the molecular structure of VD9 (Table 1). This revealed a highly oligomeric structure that contrasted with all of the hundreds of antibody structures solved to date, where a simple dimer is almost always observed.

The asymmetric unit or smallest repeating oligomer of VD9 consists of 15 molecules assembled into a close-packed left-handed helix (Figure 3(a)), and these 15-molecule helical oligomers are themselves stacked end-to-end to create a repeating fibre. The adjacent copies are related by 2-fold rotational symmetry orthogonal to the fibre axis (Figure 3(a) and (b)). There are six VD9 molecules per turn, each rotated by 60°, to give a pitch of ~35 Å or 1 VD9 domain (Figure 2(b)). All VD9 molecules are natively folded and pack antiparallel down the helix predominantly at 45° with respect to the fibre axis; the inter-subunit interactions create extended β-sheets across antiparallel packed monomers (Figure 3(c)). The centre of the fibre is hollow, comprising a water-filled nanotube of 20 Å diameter (Figure 3(d)), and with an overall diameter (about 70 Å), which is similar to that of VD9 aggregate fibres grown in solution and visualised by electron microscopy. The pentadecamer is stabilised by two novel interfaces, the first involving inter-digitation of hydrophobic side-chains and the second main-chain β-zipper interactions (Figure 4).

The first interface involves residues from and adjacent to the three complementarity- determing region (CDR) loops. Six tyrosine residues at positions 32, 49 and 92 on each molecule form a herringbone packing interaction consisting of π–π or edge-to-plane stacking of the aromatic rings (Figure 4(a)). The OH groups of tyrosine residues 32 and 92 are orientated to allow hydrogen bonding with opposing asparagine and glutamine side-chains (such that Tyr32A interacts with Asn34B and Tyr92A with Gln55B). Finally, there is a hydrophobic stacking interaction between Tyr92 and the planar side-chain of Asp49. These interface residues are remarkably conserved amongst Vκ sequences. The three tyrosine residues (32, 49 and 92) are found in 67%, 91% and 22% of Vκ sequences, respectively, whilst Asn34 and Gln55 are present in 20% and 35%, respectively. Furthermore, the VD9 residue is generally the most prevalent, suggesting that this interface could form in many Vκ antibodies.

The second interface is formed by the association of antiparallel β-edge strands (residues 9–12) from adjacent monomers to create an inter-molecular ten-stranded β-sheet (Figures 2(c) and 3(b)). The edge strands interact through a classic β-zipper main-chain hydrogen bond network, involving six close-paired hydrogen bonds between residues 9–12 (SSLS) (Figure 4(b)). The resulting inter-molecular β-sheet has all the characteristics of a standard antiparallel architecture. The $C^{\alpha}$ atoms of the edge strands align both with themselves and with the $C^{\alpha}$ atoms of the additional intra-molecular strands. Furthermore, the side-chain periodicity is the same across the sheet such that adjacent residues are all above or all below the plane of the sheet. The resulting "pleating" creates a typically shortened $C^{\alpha}$–$C^{\alpha}{}^{+2}$ atom distance of between 6–7 Å and a regular "sideways" distance between adjacent $C^{\alpha}$ atoms across the sheet of ~5 Å. Finally, this antiparallel rather than parallel packing of adjacent edge strands creates a strong oligomerisation interface as it allows the amide groups (and thus the inter-strand hydrogen bonds) to be planar and energetically most favourable. There are further packing interactions behind the zipper, including the side-chains of Val19, Pro8 and Thr20 (which create a small hydrophobic core) and hydrogen bonds from Ser7 to Thr20 and Thr22 (Figure 4(c)). Together, these interactions create a solvent-excluded interface of 790 Å.

When we superposed the light-chain from the solved Protein L–Fv complex[4] with one of the molecules from the VD9 oligomer, we found that the four β-strands of Protein L and the first four β-strands of the next VκD molecule in the oligomer came into close alignment. Indeed, Protein L binding and VD9 oligomerisation forms an identical hydrogen bond network with the same residues (Figure 4(d)). The distances and angles of the hydrogen bonds deviate, respectively, by less than 0.2 Å and 6°. The only difference is that the VκD structure appears to have an additional hydrogen bond between the amide carbonyl and nitrogen atom of residues 12 and 9, respectively (in the corresponding residue, 835, in Protein L the nitrogen points inwards).

To confirm the role of this interface in oligomerisation and Protein L binding we made a series of proline mutants to disrupt the β-edge. Oligomerisation was assessed by comparing aggregation at 4 mg/ml, and Protein L binding by capture on Protein L–agarose. The introduction of a single proline at residue 12 disrupted both oligomerisation and Protein L binding. Furthermore, this had a dramatic affect on both the solubility and expression of VD9, the mutant not aggregating at concentrations >25 mg/ml.

Taken together our results suggest that the exposed β-edge at residues 9–12 facilitates both aggregation and binding of super-antigen. Despite

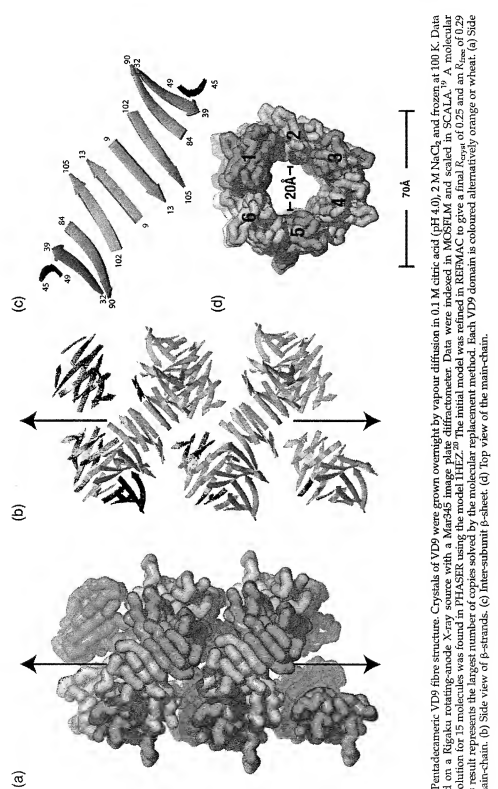



**Figure 3.** Pentadecameric VD9 fibre structure. Crystals of VD9 were grown overnight by vapour diffusion in 0.1 M citric acid (pH 4.0), 2 M $NaCl_2$ and frozen at 100 K. Data were collected on a Rigaku rotating-anode X-ray source with a Mar345 image plate diffractometer. Data were indexed in MOSFLM and scaled in SCALA.[19] A molecular replacement solution for 15 molecules was found in PHASER using the model 1HEZ.[20] The initial model was refined in REFMAC to give a final $R_{cryst}$ of 0.25 and an $R_{free}$ of 0.29 (Table 1). This result represents the largest number of copies solved by the molecular replacement method. Each VD9 domain is coloured alternatively orange or wheat. (a) Side view of the main-chain. (b) Side view of β-strands. (c) Inter-subunit β-sheet. (d) Top view of the main-chain.

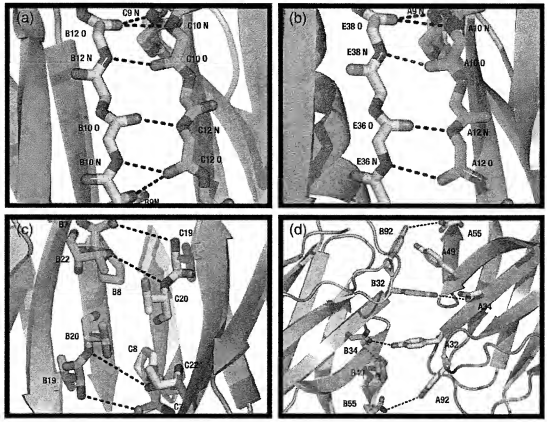



**Figure 4.** VD9 fibre interface interactions. Opposing monomers are coloured orange or wheat. Atoms are labelled with chain identifier, residue number and atom type in that order. Putative hydrogen bonds are indicated with a broken line. (a) Tyrosine hydrophobic-stacking interface. (b) β-Zipper interface, front view. (c) β-zipper interface, back view. (d) Binding interface between Protein L (orange) and Vκ (wheat).

these pathogenic effects, the design of the Vκ strand switch is highly conserved. From an earlier analysis of switch variants,[7] the wild-type appeared to be the most energetically stable, but these variants also proved more prone to aggregation during bacterial expression. This is entirely consistent with our findings for VD9, which is both aggregation-prone and thermodynamically stable with a $\Delta G$ value of 9.9 kcal mol$^{-1}$ and a $T_m$ value of 66.6 °C (using 5 μM VD9 in phosphate-buffered saline (PBS) (pH 7.4) at 25 °C and 85 °C and recorded at 235 nm. The unfolding curves were assumed to be two-state and fitted as described[8] using a $\Delta C_p$ contribution of 12 cal per amino acid residue to give $T_m$ and $\Delta G_{N\text{-}U}$; data not shown). We suggest the two strand switch designs each have their own advantages and disadvantages. Thus, in the κ design a short segment of β-strand is exposed, leading to super-antigen interaction and association with light chain deposition diseases,[2,3] whereas in the λ design the strands are capped but the domain is less stable (and associated with amyloidosis[1]). In turn this suggests that light chain deposition diseases may be driven by native-like interactions between thermodynamically stable domains, in contrast to light chain amyloidosis, which seems more likely to be driven by the interactions of unfolded polypeptide.[9,10]

Both the β-edge and CDR interfaces observed here between free κ-domains should also be permitted between free κ-light chains or between Bence–Jones κ-light chain dimers *in vivo*. By contrast, inspection of three-dimensional models of IgG[11] indicates that due to steric constraints, only one or other set of interactions is permitted in intact IgG. Light-chain deposition disease is characterised by abnormal levels of free κ-chain synthesis, either as a result of lymphoproliferative disorders such as myeloma, which account for 60% of cases,[12] or by unbalanced immunoglobulin synthesis in bone marrow cells,[13] and biopsies of diseased glomerular membranes stain for κ-chains but not IgG or λ.[14] Whether *in vivo* deposits combine the interfaces as seen in the crystal is unclear. Crystalline κ-chain deposits have been observed in glomerulonephritis: κ-chain crystal deposition in proximal tubular cells is a cause of proximal tubulopathy[15] whilst cytoplasmic crystalline κ-chain inclusions have been observed in malignant plasma cells in Fanconi's syndrome.[16]

It is also possible that the exposed segment of the Vκ domain confers a functional benefit, for example in pathogen recognition. There are still many aspects

of the antibody immune response which are poorly understood, particularly how the primary response, with its diversity limited to the number of circulating B cells, can bind with physiological affinity to an almost infinite diversity of pathogen molecules. The observation that polyclonal antibodies are often more effective at neutralising antigens than individual constituent antibodies has led to the suggestion that they operate synergistically. Potentially, the Vκ strand switch sequence could allow antibodies from a polyclonal response to multimerise through edge–edge interaction, significantly increasing affinity through avidity. Calarese *et al.* recently showed that variable domains can adopt unusual quaternary structures. In their case, an anti HIV-1 antibody underwent heavy chain domain exchange in order to bind a repetitive carbohydrate epitope.[17]

β-Edges have long been hypothesised to play a central role in protein aggregation. Indeed, it has been suggested that the protection of edges may well have been the primary driving force behind present day β-protein topology.[5,18] Our results show that in antibodies the failure to protect a short β-edge in Vκ domains is exploited by pathogens and may also lead to light chain deposition disease. Subtle differences in strand protection in Vκ and Vλ domains therefore appear to have major consequences for pathology.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jmb.2006.10.093

## References


1. Bellotti, V., Mangione, P. & Merlini, G. (2000). Review: immunoglobulin light chain amyloidosis–the archetype of structural and pathogenic variability. *J. Struct. Biol.* 130, 280–289.
2. Deveney, L., Deret, S. & Aucouturier, P. (1994). Over-representation of V kappa IV subgroup in light chain deposition disease. *Immunol. Letters*, 42, 63–66.
3. Myatt, E. A., Westholm, F. A., Weiss, D. T., Solomon, A., Schiffer, M. & Stevens, F. J. (1994). Pathogenic potential of human monoclonal immunoglobulin light chains: relationship of *in vitro* aggregation to *in vivo* organ deposition. *Proc. Natl Acad. Sci. USA*, 91, 3034–3038.
4. Graille, M., Stura, E. A., Housden, N. G., Beckingham, J. A., Bottomley, S. P., Beale, D. *et al.* (2001). Complex between *Peptostreptococcus magnus* protein L and a human antibody reveals structural convergence in the interaction modes of Fab binding proteins. *Structure (Camb)*, 9, 679–687.
5. Richardson, J. S. & Richardson, D. C. (2002). Natural beta-sheet proteins use negative design to avoid edge-to-edge aggregation. *Proc. Natl Acad. Sci. USA*, 99, 2754–2759.
6. Cox, J. P., Tomlinson, I. M. & Winter, G. (1994). A directory of human germ-line V kappa segments reveals a strong bias in their usage. *Eur. J. Immunol.* 24, 827–836.
7. Spada, S., Honegger, A. & Pluckthun, A. (1998). Reproducing the natural evolution of protein structural features with the selectively infective phage (SIP) technology. The kink in the first strand of antibody kappa domains. *J. Mol. Biol.* 283, 395–407.
8. Pace, C. N. S., J. M. (1997). In *Protein Structure, A Practical Approach* (Creighton, T. E., ed), 2nd edit., vol. 1, Oxford University Press, New York.
9. Ferguson, N., Berriman, J., Petrovich, M., Sharpe, T. D., Finch, J. T. & Fersht, A. R. (2003). Rapid amyloid fiber formation from the fast-folding WW domain FBP28. *Proc. Natl Acad. Sci. USA*, 100, 9814–9819.
10. Uversky, V. N. & Fink, A. L. (2004). Conformational constraints for amyloid fibrillation: the importance of being unfolded. *Biochim. Biophys. Acta*, 1698, 131–153.
11. Harris, L. J., Skaletsky, E. & McPherson, A. (1998). Crystallographic structure of an intact IgG1 monoclonal antibody. *J. Mol. Biol.* 275, 861–872.
12. Tanenbaum, N. D., Howell, D. N., Middleton, J. P. & Spurney, R. F. (2005). Lambda light chain deposition disease in a renal allograft. *Transplant Proc.* 37, 4289–4292.
13. Gallo, G. R., Feiner, H. D., Katz, L. A., Feldman, G. M., Correa, E. B., Chuba, J. V. & Buxbaum, J. N. (1980). Nodular glomerulopathy associated with nonamyloidotic kappa light chain deposits and excess immunoglobulin light chain synthesis. *Am. J. Pathol.* 99, 621–644.
14. Chang, A., Peutz-Kootstra, C. J., Richardson, C. A. & Alpers, C. E. (2005). Expanding the pathologic spectrum of light chain deposition disease: a rare variant with clinical follow-up of 7 years. *Mod. Pathol.* 18, 998–1004.
15. Cai, G., Sidhu, G. S., Wieczorek, R., Gu, X., Herrera, G. A., Cubukcu-Dimopulo, O. & Kahn, T. (2006). Plasma cell dyscrasia with kappa light-chain crystals in proximal tubular cells: a histological, immunofluorescent, and ultrastructural study. *Ultrastruct. Pathol.* 30, 315–319.
16. Bridoux, F., Sirac, C., Hugue, V., Decourt, C., Thierry, A., Quellard, N. *et al.* (2005). Fanconi's syndrome induced by a monoclonal Vkappa3 light chain in Waldenstrom's macroglobulinemia. *Am. J. Kidney Dis.* 45, 749–757.
17. Calarese, D. A., Scanlan, C. N., Zwick, M. B., Deechongkit, S., Mimura, Y., Kunert, R. *et al.* (2003). Antibody domain exchange is an immunological solution to carbohydrate cluster recognition. *Science*, 300, 2065–2071.
18. Siepen, J. A., Radford, S. E. & Westhead, D. R. (2003). Beta edge strands in protein structure prediction and aggregation. *Protein Sci.* 12, 2348–2359.
19. CCP4. (1994). Collaborative Computational Project 4. *Acta Crystallogr. sect. D*, 50, 760–763.
20. Storoni, L. C., McCoy, A. J. & Read, R. J. (2004). Likelihood-enhanced fast rotation functions. *Acta Crystallog. sect. D*, 60, 432–438.
21. Tennent, G. (1999). Isolation and characterization of amyloid fibrils from tissue. In *Methods in Enzymology* (Wetzel, R., ed), p. 309, Academic Press Ltd, San Diego, CA.


*Edited by I. Wilson*

*(Received 12 September 2006; received in revised form 25 October 2006; accepted 26 October 2006)*
Available online 3 November 2006

Structure, Vol. 9, 679–687, August, 2001, ©2001 Elsevier Science Ltd. All rights reserved. PII S0969-2126(01)00630-X

# Complex between *Peptostreptococcus magnus* Protein L and a Human Antibody Reveals Structural Convergence in the Interaction Modes of Fab Binding Proteins

Marc Graille,[1] Enrico A. Stura,[1,5] Nicholas G. Housden,[2] Jennifer A. Beckingham,[2] Stephen P. Bottomley,[2] Dennis Beale,[3] Michael J. Taussig,[3] Brian J. Sutton,[4] Michael G. Gore,[2] and Jean-Baptiste Charbonnier[1,5]
[1]Département d'Ingénierie et d'Etudes des Protéines
Commissariat à l'Energie Atomique
Centre d'Etudes Saclay
Gif-sur-Yvette F-91191
France
[2]Department of Biochemistry
Institute of Biomolecular Sciences
University of Southampton
Bassett Crescent East
Southampton SO16 7PX
United Kingdom
[3]Laboratory of Molecular Recognition
The Babraham Institute
Babraham, Cambridge CB2 4AT
United Kingdom
[4]The Randall Centre
King's College London
Guy's Campus
London SE1 1UL
United Kingdom

## Summary

**Background:** *Peptostreptococcus magnus* protein L (PpL) is a multidomain, bacterial surface protein whose presence correlates with virulence. It consists of up to five homologous immunoglobulin binding domains that interact with the variable ($V_L$) regions of kappa light chains found on two thirds of mammalian antibodies.

**Results:** We refined the crystal structure of the complex between a human antibody Fab fragment (2A2) and a single PpL domain (61 residues) to 2.7 Å. The asymmetric unit contains two Fab molecules sandwiching a single PpL domain, which contacts similar $V_L$ framework regions of two light chains via independent interfaces. The residues contacted on $V_L$ are remote from the hypervariable loops. One PpL-$V_K$ interface agrees with previous biochemical data, while the second is novel. Site-directed mutagenesis and analytical-centrifugation studies suggest that the two PpL binding sites have markedly different affinities for $V_L$. The PpL residues in both interactions are well conserved among different *Peptostreptococcus magnus* strains. The Fab contact positions identified in the complex explain the high specificity of PpL for antibodies with kappa rather than lambda chains.

**Conclusions:** The PpL-Fab complex shows the first interaction of a bacterial virulence factor with a Fab light chain outside the conventional combining site. Structural comparison with two other bacterial proteins interacting with the Fab heavy chain shows that PpL, structurally homologous to streptococcal SpG domains, shares with the latter a similar binding mode. These two bacterial surface proteins interact with their respective immunoglobulin regions through a similar β zipper interaction.

[5]Correspondence: estura@cea.fr (E.A.S.)
jb.charbonnier@cea.fr (J.B.C.)

## Introduction

*Peptostreptococcus magnus* protein L (PpL: whole protein L. PpL domain: individual domain of protein L) is a cell wall-anchored protein able to interact with a large repertoire of mammalian immunoglobulins (Ig) [1]. *Staphylococcus aureus* protein A (SpA) [2] and streptococcal protein G (SpG) [3] also share this property, although they recognize different Ig regions. PpL is present at the surface of about 10% of *Peptostreptococcus magnus* strains [4]. It is a 76–106 kDa protein containing four or five highly homologous, consecutive extracellular Ig binding domains (depending on the bacterial strain from which it is isolated [5, 6]). The structure of PpL domain B, (76 amino acids) has been determined by NMR spectroscopy [7]. The fold of this domain is similar to that of the SpG Ig binding domains. It consists of a β sheet composed of two pairs of anti-parallel β strands and an α helix that lies on top of the sheet.

PpL interacts with Ig light chains [1], notably with the kappa light-chain variable region ($V_L$) from humans and other mammals [8, 9]. When present on the bacterial surface, PpL has been described as a virulence factor of bacterial vaginosis in different clinical specimens [4]. It has also been shown that PpL induces histamine release by basophils and mast cells, presumably by cross-linking IgE bound to $F_{C\varepsilon}$ receptors [10, 11].

PpL and SpA single domains are targets for protein engineering due to their ability to bind the variable regions (Fv) of a large population of antibodies. This unique property makes them valuable tools in biotechnology for the purification and recognition of recombinant single-chain (sc) Fv. Such engineered antibodies are increasingly used in the optimization of specificity and affinity by phage display and other in vitro evolutionary mutagenesis techniques [12]. The structural characterization of PpL and SpA binding properties is useful for defining the spectrum of Fvs, which bind to these domains, based on their sequences.

We report in this study the crystal structure of the human antibody Fab 2A2 complexed through its $V_L$ region to a PpL domain. The same Fab was the subject of a previous crystallographic study describing its complex with a *Staphylococcus aureus* protein A (SpA) domain, which binds to the antibody $V_H$ region [13]. Here, we compare the location of these two Ig binding domains on either side of the Fv region of Fab 2A2 in relation to the antigen-combining site. Unexpectedly, we find that

**Key words:** immunoglobulin binding proteins (IBP); antibody complex; cell surface protein; X-ray structure

there are two PpL-Fab interfaces, such that one PpL domain has two separate regions that can interact with kappa light chains and is capable of binding two Fabs simultaneously. The two interfaces involve similar sites on the $V_L$ domains. One of the PpL–Fab interfaces conforms to previous biochemical data, while the second is novel. Site-directed mutagenesis and analytical-centrifugation studies suggest that the two PpL combining regions have markedly different affinities for $V_L$. The high specificity of PpL for kappa (the largest mammalian $V_L$ gene family), rather than lambda chains, is discussed in light of the structure of the complex. We analyze the sequence diversity of the PpL domains at positions involved in the interaction with $V_L$ regions and discuss avidity effects reported between whole PpL and Ig. Furthermore, by comparing the interactions of PpL and SpG with $V_L$ and $C_H1$, respectively, we note that at a structural level, the Fab binding modes of these two bacterial proteins show a degree of convergence.

## Results and Discussion

### The First $V_L$–PpL Interface

The asymmetric unit has one PpL domain and two Fab molecules. The PpL domain is in close contact with the $V_L$ region from both Fabs (Figure 1a). This stoichiometry was unexpected because no previous studies raised any evidence for one PpL domain being able to complex to two $V_L$ regions simultaneously [6, 14].

The first interface buries a total solvent-accessible area of 1300 Å$^2$ with approximately equal contributions from both molecules, as determined with a 1.4 Å radius probe, and is remote from the Ig heavy chain. There is no conformational change in the backbone of either partner upon binding. The PpL domain C* used in this study has 94% sequence identity with the C4 domain [15]. Its affinity for kappa chains is comparable (130 nM) to the one measured for PpL domain B1 [14, 15]. Domain C* maintains the same three-dimensional structure as PpL domain B1 determined by NMR [7], and the two structures superpose with an rmsd of 1.31 Å over 59 residues. The complexed PpL domain C* superposes with an rmsd of 0.39 Å over 61 residues of the crystallographic structure of free PpL domain C* (B.J.S., unpublished data). Similarly, the variable region of the complexed Fabs superimposes with an rmsd of 0.8 Å over 215 residues with the free structure determined previously [13].

The interaction with PpL involves 13 residues from the Fab (Figures 1b and 1c). Ten are located in framework region 1 (FR1). The others are Lys-L107 from the segment connecting the $V_L$ to the $C_L$ region, Glu-L143 from the $C_L$ region, and Arg-L24 from the CDR-L1 region [16] on strand B. However, this residue does not belong to the $V_L$ hypervariable loops according to the structural definition of Chothia [17] or to the positions frequently identified by the contact with antigens [18]. Compared with SpA, which binds to the variable region of the same human antibody (Fab 2A2) [13], PpL is farther away from the center of the antigen binding site (23 Å compared to 16 Å). Hence, like SpA, PpL binding should not affect accessibility to the antigen-combining site, as also suggested by competition assays [19].

Of the amino acids in the PpL domain, 12, located mainly on strand β2 and the α helix, are involved in this interaction (Figure 2). This interface is characterized by six hydrogen bonds (Figure 1b, Table 1). Three are between main-chain atoms located on PpL strand β2 and on strand A from the first Fab; they thus join the β sheets of the Fab and the protein L into a unique sheet through a β zipper interaction.

Heteronuclear NMR spactroscopy has been used for mapping backbone positions of the PpL domain B1 [20] involved in the interaction with the $V_L$ region. Most positions identified for domain B1 are also implicated in the first interface of domain C* described in the present study. These backbone positions are on strand β2 and the α helix of the PpL domains. A minor discrepancy with the NMR results concerns the loop between the α helix and strand β3, which does not make any contact with the Fab $V_L$ region in the crystal structure of the complex. This loop is poorly defined in the NMR structure, so the discrepancy can be attributed to a change in mobility upon complexation.

The first interface was subjected to site-directed mutagenesis of PpL domain C*. A 23-fold drop in affinity results from a Y53F substitution [21] as measured by competitive ELISA. This decrease following the loss of the tyrosine hydroxyl is as expected on the basis of the energy of a neutral hydrogen bond [22], consistent with the loss of the interaction between the Tyr-53 hydroxyl group and the Thr-L20 carbonyl group of the $V_L$ region (Table 1). Thus, the first interface explains the existing data well.

### The Second $V_L$–PpL Interface

In the crystal, the single PpL domain is sandwiched between the $V_L$ regions of two 2A2 Fabs present in the asymmetric unit (Figure 1a). The second interaction buries a total solvent-accessible area of 1400 Å$^2$, comparable to other protein-protein complexes [23]. The PpL domain C* has a different orientation in the two interactions relative to the Fab β sheet. This second interaction involves 15 $V_L$ residues, located mainly on the β strands A and B (as in the first interface) with some participation of strands D and E (Figure 1c). Out of the 15 residues from the $V_L$ region involved in this second interaction, 10 are common to the first one. On the contrary, none of the PpL residues that contribute significantly to this interface are involved in the first one. The 14 amino acids from the PpL domain involved in this second interaction come mainly from strand β3 and from the α helix (Figure 2). Six hydrogen bonds and two salt bridges mediate this interaction (Table 2), as compared to only six hydrogen bonds for the first interface (Table 1). Although unexpected from biochemical studies, this second PpL–Fab interface buries an area too large to be a crystal contact. Given that this interface buries a surface larger than 1400 Å$^2$, the probability that it is just a crystal contact can be evaluated as only 2% [24]. Thus, we believe that this interface has to be given serious attention.

What is the strength of this second interaction? At present, no definite answer can be given, but we have used mutagenesis to probe the relative contributions of these two interfaces. First, we constructed the Y64W

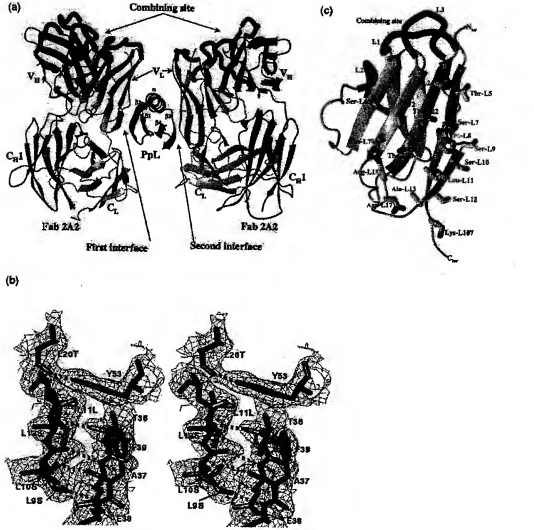



Figure 1. PpL domain C* Complexed with Human IgM Fab 2A2

(a) Ribbon representation of the 2 Fab:1 PpL domain complex present in the asymmetric unit. The PpL domain C* (red) is sandwiched between two Fabs (blue and green). Light colors represent the light chains, and dark colors represent the heavy chains. Magenta highlights the CDR loops, as defined by Chothia [17], and positions the Fab-PpL interfaces relative to the combining site. Pseudo 2-fold axis symmetry relates the two Fabs, but both interfaces are not symmetrical.

(b) Stereoview of the $2F_o - F_c$ electron density map contoured at 1$\sigma$ at the first $V_L$-PpL interface. The letter L precedes amino acids from the Fab light chain. Green dotted lines depict the three hydrogen bonds between Fab and PpL main-chain atoms. The three other hydrogen bonds are shown in magenta and involve at least one side-chain atom.

(c) Ribbon representation of the $V_L$ region of Fab that interacts with the PpL domain. Ten residues common to the first and second interfaces are in yellow. Positions in pink are only involved in the first interface, and those in light green are implicated only in the second interface. All figures were generated with MOLMOL [42] or TURBO-FRODO [43].

mutant of a PpL domain C* to have an efficient fluorescent probe so as to measure Kd by the stopped-flow method [21]. Second, we constructed two double mutants, Y53F-Y64W [21] and D55A-Y64W (this study), with the purpose of weakening, respectively, the first and second interfaces. In the Y53F-Y64W mutant, the replacement of Tyr-53 by Phe disrupts the hydrogen bond between the tyrosine hydroxyl and the L20 carbonyl in the first interface. Similarly, we chose the D55A-Y64W mutant in order to disrupt the salt bridge between Asp-

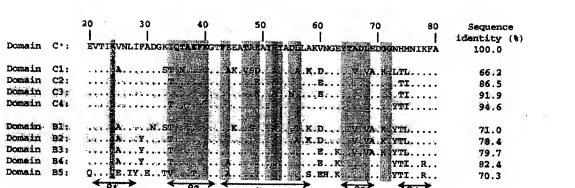



Figure 2. Alignment of the Amino Acid Sequences of the Different Ig Binding Domains from PpL

The PpL residues represented are only those visible in the crystal structure. Domain C* is the one used in this study. Domains $C_1$ to $C_4$ are from $PpL_{BRE}$ strain [6], and domains $B_1$ to $B_5$ are from $PpL_{HAR}$ strain [6]. PpL residues involved in the first interface with the Fab 2A2 are highlighted in pink, and those involved in the second binding site are in green. With the exception of Arg-52 (yellow), there is no overlap between the residues involved in each interaction. Core residues are in bold. Those involved in the second interface are mainly located on the α helix and strand β3 and are well conserved among the different PpL domains. Of note is that one mutation (Gly-71 to Lys) may reinforce or weaken this second interface. The amino acid sequence of SpG domain III is also shown with the residues involved in the interaction with the Fab shown in pink, those interacting with the $F_{C\gamma}$ in cyan, and those implicated in both interactions in yellow.

55 and Arg-L24 as well as the hydrogen bond with the Ser-L7 side chain in the second interface. These three single PpL domain constructs are correctly folded, as indicated by far-UV CD spectroscopy (data not shown). While the loss of a single hydrogen bond at the first interface has a significant effect on the dissociation constant of the $V_L$-PpL domain C* complex, the disruption of the salt bridge and the hydrogen bond mediated by Asp-56 of PpL at the second interface does not alter the Kd of the complex (Table 3). This suggests that the binding of single PpL domain C* is dominated by the first interface and that the dissociation constant of the second is at least one order of magnitude larger than the first one.

Evidence supporting the existence on a single PpL domain of two Fab-combining regions, differing in affinities by about two orders of magnitude, has been obtained recently by analytical-centrifugation studies (B.J.S. and R. Beavil, unpublished data).

This observation is reminiscent of the 1:2 stoichiometry described for the interaction between human growth hormone and its receptor [25], where two different faces of the hormone contact similar surfaces on two receptors, with the result being receptor activation [26]. By analogy, in our study a single PpL domain contacts similar surfaces on two $V_L$ regions and could bridge two Igs anchored at the membrane of B cells. The biological relevance of this second interface and the potential mitogenic activity of a single PpL domain on B cells are under investigation.

Table 1. Hydrogen Bonds Involved in the First $V_L$-PpL Interface

| $V_L$ Region | PpL Domain C* | Distance (Å) |
|---|---|---|
| Ser L-9 N | O Glu 38 | 2.8 |
| Ser L-9 $O_\gamma$ | N 40 Lys | 3.1 |
| Ser L-10 $O_\gamma$ | $O_{\varepsilon 1}$ Glu 38 | 2.9 |
| Ser L-10 O | N Glu 38 | 3.1 |
| Ser L-12 N | O Thr 36 | 2.7 |
| Thr L-20 O | $O_\eta$ Tyr 53 | 2.6 |

Hydrogen bond distances were calculated with the program CONTACTS [44] with a maximum distance cutoff of 3.4 Å and defined according to the criteria of McDonald and Thornton [45].

## $V_L$ Specificity of the PpL Binding Interaction

PpL has been reported to bind efficiently to $V_L$ regions of about two thirds of the human Ig repertoire and to thus encompass mainly the κ1, κ3, and κ4 subgroups [9] but neither κ2 nor any λ subgroups. The PpL kappa specificity observed for human antibodies extends to various other mammalian species [9]. To analyze the structural determinants of this specificity, we have superimposed on the 2A2 $V_L$ region the equivalent structures found on different kappa and lambda subgroups. We observed that PpL binding ability is mostly concen-

Table 2. Hydrogen Bonds and Salt Bridges Involved in the Second $V_L$-PpL Interface

| $V_L$ Region | PpL Domain C* | Distance (Å) |
|---|---|---|
| Ser L-7 $O_\gamma$ | Oδ1 Asp 56 | 2.9 |
| Ser L-10 $O_\gamma$ | $O_{\gamma 1}$ Thr 65 | 3.2 |
| Ser L-12 N | O Ala 66 | 2.9 |
| Ser L-12 O | N Leu 68 | 2.6 |
| Arg L-18 N | O Gly 71 | 3.1 |
| Arg L-24 $N_{\eta 1}$ | Oδ1 Asp 55 | 2.7 |
| Lys L-107 $N_\zeta$ | Oδ1 Asp 67 | 2.6 |
| Lys L-107 $N_\zeta$ | O Leu 68 | 2.9 |

Hydrogen bond and salt bridge distances were calculated with the program CONTACTS [44] with a maximum distance cutoff of 3.4 Å and defined according to the criteria of McDonald and Thornton [45].

Table 3. Dissociation Constants between PpL Single Domain C* and Human κ Chain as Determined by Stopped-Flow Fluorescence Measurements

| PpL Construct | Kd ($\mu$M) | Comment |
|---|---|---|
| Wild type | 0.13[1] | |
| Y64 W | 0.17[2] | Fluorescent probe |
| Y53P–Y64W | 2.9[2] | Hydrogen bond on interface 1 disrupted |
| D65A–Y64W | 0.145 | Hydrogen bond and salt bridge on interface 2 disrupted |

[1]From Beckingham et al [14].
[2]From Beckingham et al [21].

trated on the L5 to L12 segment of the $V_L$ region and that the interaction is dependent on the main-chain conformation of this segment, on which conformation the β zipper interaction also depends. Particularly important are residues L8–L12, which bury more than 80% of their total accessible area. The $V_L$ residues of this segment and at other positions of the first interface are well conserved among the recognized $V_L$ regions (Table 4). The main chain superimposes well onto $V_L$ structures of the κ1, κ3, and κ4 subgroups (rmsd of 0.3–0.5 Å over the eight residues from L5 to L12). Most κ2 subgroup sequences have a proline residue at position L12 that introduces major steric hindrance by pointing the Pro ring toward the PpL domain. For the λ subgroup, two observations account well for the absent or weak binding activity found for these chains. Firstly, the L8–L12 segment is one residue shorter, resulting in backbone conformational differences, which may alter the β zipper interaction (rmsd of 1–1.2 Å over the seven residues from L5 to L12). Secondly, due to the reduced size of this segment, a cavity is created around positions L7 and L8. The Fab recognition mode developed by the PpL domain is highly dependent on the backbone conformation, different from that of SpA, which relies more on a side chain-specific recognition mode [13].

### Avidity of Whole PpL for Immunoglobulins

PpL contains four or five highly homologous, consecutive extracellular Ig binding domains in tandem, depending on the strain. At what level are the positions involved in the first interface conserved between the different domains? Sequence alignment of the four C domains and five B domains with the C* domain (Figure 2) shows that domain C1, which differs at 20 positions out of 61, is the most distant from domain C*. The differences on C1 are distributed all over the domain, with higher incidence on the α helix and strands β2, β3, and β4. The other C domains have a higher degree of sequence identity with C*. The B domains have from 10 to 17 well-distributed changes compared to C*.

Structurally, a core of seven critical residues can be identified. These residues are those largely buried upon complex formation or those involved in hydrogen bonds: Gln-35 to Lys-40 from strand β2 and Tyr-53 from helix α. These seven residues are strictly conserved in eight out of the ten PpL domains (Figure 2). Domains C1 and B5 each have only one nondisruptive difference. The changes from Thr-36 to Asn in C1 and from Glu-38 to Thr in B5 could weaken the interaction but not disrupt it. Residue changes outside the structural core in domains C2 and C3 could result in a slightly lower affinity for κ chains compared to that of domain C*. Residues 49 and 52 from domain C* are, respectively, Glu and Arg, which make an internal salt bridge in the middle of the α helix. In domains C2 and C3, these residues are replaced, respectively, by Lys and Ala. This disrupts the salt bridge and places the Lys in close proximity to two Arg residues, and this further results in an unfavorable accumulation of positive charges.

As is the case for the five SpA domains with respect to their interaction with $V_H$ region [13], most PpL domains should also conserve all their hydrogen bonding interactions. Thus, we would expect that all the PpL domains would bind to the $V_L$ region in a similar way so that a whole PpL would contain at least four Ig binding sites. The equivalence in binding between the domains is consistent with previous studies on the binding properties of different PpL constructs. These studies show that avidity increases with the number of Ig binding domains [6]. In fact, a four-domain PpL has an avidity 100-fold higher than the single domain, and a fifth Ig binding domain does not improve avidity further [6]. By superimposing a PpL–Fab complex on each $V_L$ region of whole IgG [27, 28], we can show that the distance between the C terminus of a PpL domain bound to one Fab and the N terminus of the other bound PpL ranges from 60–120 Å depending on the Fab hinge angle. Four domains (including their interdomain linkers of 16 amino

Table 4. Sequences of the Human $V_L$ Gene Families at the Positions Involved in the First Interface with PpL

| $V_L$ Subgroup | PpL Binding | L5 | *L6* | *L7* | L8 | L9 | L10 | L11 | L12 | L13 | L18 | L20 | L22 | L24 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| κ1 (19) | + | T | *Q* | *S* | P | S | S | L | S | A | R | T | T | R |
| κ3 (7) | + | T | Q | S | P | A | T | L | S | L,V | R | T | S | R,G |
| κ4 (1) | + | T | Q | *S* | P | D | S | L | A | V | R | T | N | K |
| κ2 (9) | – | T | Q | *S* | P | L | S | L | P,S | V | P | S | S | R,K |
| λ1 (5) | – | T | Q | * | P | P | S | V,A | S | G,A,E | R,K | T | S | S,T |
| λ2 (5) | – | T | Q | * | P | P,R,A | S | V,A | S | G | S | T | S | T |
| λ3 (9) | – | T | Q | * | P | P | S,A | V | S | V | T,M | R | T | S,G,Q |
| λ4 (3) | – | T | Q | * | P,S | P,S | S | A | S | A | S | K | T | T |
| λ5 (3) | – | T | Q | * | P | P,A,S | S | L,S,H | S | A | S | R,S | T | T,M |

The numbers in parentheses for each $V_L$ group indicate the numbers of human germ-line sequences reported in the V Base database (Medical Research Council, Centre for Protein Engineering; http://www.mrc-cpe.cam.ac.uk/imt-doc/). Families that represent more than 10% of the total genes are reported. Amino acids observed in fewer than 15% of functional germline human genes are not reported. Positions 6 and 7 are shown in italics, although they are not in direct contact with PpL to aid the structural alignment between λ and κ chains. The asterisks (*) indicate missing residues. The PpL binding properties are taken from Nilson *et al.* [8].

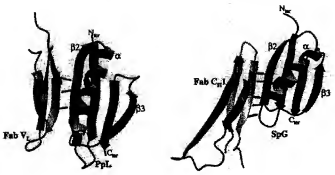



Figure 3. Comparison of SpG-$C_{\kappa}1$ and PpL-$V_L$ interactions

Orienting the SpG and PpL β sheets in an analogous manner underscores the structural similarities of both interactions. Strand β2 of both domains interacts with, respectively, an Fab $V_L$ (blue) or an Fab $C_{\kappa}1$ (green) β strand (colored lines represent hydrogen bonds between main-chain atoms). In both cases, the bacterial strand β2 forms a continuous β sheet with the Fab through a β zipper interaction.

acids each) could span this distance, and this could explain the strong avidity effect observed on Ig binding. The same pattern of modular domains binding to Igs is shared by SpA and SpG, which like PpL seem to have evolved high avidity for Ig by cassette duplication to allow multiple attachments [29].

Although the Kd of the second interface may be substantially higher than the first, we cannot exclude the possibility that the second interface could be structurally important for PpL avidity. A PpL molecule bound to Ig through the high-affinity site of one domain could use the low-affinity site from another domain to interact with the Ig molecules when this would be sterically more favorable than interacting via the first interface.

## Structural Convergence between the Fab Binding Modes of SpG and PpL

*Streptococci* and *Peptostreptococci* are found in the same habitats, including the human intestinal and genital tracts. Their cell surface proteins share many common features, and it has been proposed that a transfer of gene fragments may have occurred between these two infective bacteria [29]. With only 15% sequence identity, the Ig binding domains are the least homologous regions between these cell surface proteins. Despite the low sequence identity, these domains share a common fold, with an α helix packed against a four-stranded β sheet. The NMR and crystal structures of both domains have shown that the main structural difference is the orientation of the α helices [7, 30]. The helix runs almost parallel to the β strand direction in the PpL β sheet, whereas in SpG it runs diagonally across the sheet. This arises from a difference in the loop between the α helix and the strand β3, which is one residue shorter in PpL. Since this is the region of SpG that binds to Fc, it may in part explain the absence of Fc binding by PpL.

The interaction of SpG with $C_{\kappa}1$ and that of the first interface of PpL with $V_L$ regions of Fab (Figure 3) share similar features. Both domains bury equivalent surface areas upon binding to the Fab and have two common structural features. Firstly, the same region of these two domains is buried in the interactions since SpG also interacts through its strand β2 [31]. Secondly, in both cases this strand extends a β sheet of the Fab through a β zipper interaction. This interaction involves five main-chain/main-chain hydrogen bonds in the SpG-$C_{\kappa}1$ interface [30] but only three hydrogen bonds in the PpL-$V_L$ first interface. The external strand A of the $V_L$ region involved in the interaction presents a bulge due to proline at position L8 in the middle of the strand, and this bulge shortens the β zipper. Although the SpG and PpL domains are clearly the result of divergent evolution [29], they have maintained a common binding strategy for interacting with the β strands from different Ig domains (Figure 3) through a main-chain-to-main-chain hydrogen bonding network.

Like SpA, PpL domains target a wide repertoire of human immunoglobulins. As none of those bacterial domains interferes with antibody-antigen recognition, they

Table 5. Summary of Crystallographic Analysis

| | |
|---|---|
| **Data collection[1]** | |
| Resolution (Å) | 20–2.6 |
| Unique reflections | 29,954 |
| Redundancy | 4.66 (5.0) |
| I/σ | 11.4 (2.5) |
| $R_{sym}$ (%) | 11.6 (46.6) |
| Completeness (%) | 95.2 (93.7) |
| Space group | $P2_12_12_1$ |
| Unit cell parameter | a = 55.2 Å, b = 87.3 Å, c = 210.5 Å |
| **Molecular replacement** | |
| Resolution (Å) | 20–4.0 |
| Correlation coefficient (%) | 61.4 (46.9)[2] |
| **Refinement** | |
| Resolution (Å) | 20–2.7 |
| $R_{work}$ (%) | 21.5 |
| $R_{free}$ (%)[3] | 27.8 |
| Rmsd | |
| Bonds (Å) | 0.007 |
| Angles (°) | 1.424 |
| Averaged B factor (Å²) | 45.5 |
| Luzzati coordinates error (Å) | 0.34 |
| Number of water molecules | 112 |

$R_{work}$ and $R_{free}$ values in the last shell (2.75–2.7 Å) are, respectively, 31.3% and 34.7%.
[1]Numbers in bracket give values for the highest resolution shell (2.65–2.6 Å).
[2]Correlation coefficient for the next highest solution is indicated in bracket.
[3]Value of $R_{free}$ calculated for 5% randomly chosen reflections not included in the refinement.

may be used to aid the crystallization of antigen-Fab [32, 33] or membrane protein-Fv complexes [34]. We suggest that these small protein domains should be used in a combinatorial manner to increase the variety of possible crystal contacts that can be formed. This would help to solve some problematic crystallization cases. Knowledge of the structure of the PpL-Fab complex will facilitate the engineering of PpL variants with extended binding capacity, for example to encompass λ chains and improve their utility as reagents for Ig detection and purification.

## Biological Implications

Several pathogenic bacteria present cell surface proteins with immunoglobulin binding properties that confer on them an advantage during invasion and colonization of host tissues. Some of these bacterial proteins interact with a wide proportion of the immunoglobulin repertoire through interactions with constant regions of immunoglobulin that are highly conserved in sequence. Two bacterial proteins, *Staphylococcus aureus* protein A (SpA) and *Peptostreptococcus magnus* protein L (PpL), interact with a wide range of antibodies by targeting variable, rather than constant, regions of Fab.

We recently reported the crystal structure of an SpA domain in contact with the variable-heavy ($V_H$) region of an Fab [13] and we now report the crystal structure of a PpL domain in complex with the variable-light ($V_L$) region of the same Fab. Both bacterial domains interact with the framework part of these variable regions without contacting the hypervariable loops. The positions identified in both complexes account well for the wide repertoire of immunoglobulins recognized by both bacterial domains. In contrast with SpA, which targets conserved residues of the $V_H$ external β sheet, PpL targets a portion of the $V_L$ region that is not strictly conserved in sequence but that maintains a constant backbone conformation among most $V_L$ regions encoded by κ genes.

A single domain of this multidomain protein manifests PpL activity. The PpL domain appears to have two independent binding sites for $V_L$, which interact with very similar areas on the light chain but with markedly different affinities. The residues involved in the $V_L$ interaction are well conserved among the different PpL domains. The interaction of the single domain accounts well for two of the properties reported for the whole protein L. Firstly, PpL exhibits an avidity effect for whole immunoglobulin, and the overall position of a single bacterial domain suggests that a whole PpL with up to five domains in tandem should be able to bind the two $V_L$ regions of an IgG. Secondly, PpL could induce histamine release by bridging two IgE bound to their $F_{C\varepsilon}$ receptors at the surface of human basophils or mast cells. The conservation of PpL positions in all domains is also in agreement with this last property.

## Experimental Procedures

### Crystallization and Data Collection

The 2A2 IgM rheumatoid factor was secreted by a hybridome line created from synovial B cells of a rheumatoid arthritis patient [35]. Fab 2A2 was produced by trypsin cleavage of the IgM [35]. The $V_L$ chain of 2A2 belongs to the κ1 subtype recognized by PpL. The PpL domain C* used in the crystallization has 94% sequence identity with the C4 domain from PpL (4 mutations over 61 amino acids) and was produced by recombinant expression [15].

Crystals were grown by vapor diffusion at room temperature in sitting drops by the mixture of a reservoir solution (13%–16% [wt/wt] monomethyl polyethylene glycol (MPEG) 5000, 100 mM imidazole malate [pH 8.5]) with an equal volume of protein solution (5 mg/ml) containing different Fab 2A2-PpL molar ratios. The ratio was optimized with the aid of streak seeding [37]. Interestingly, the complex crystals grew at a pH that was at or above the upper limit for the growth of unliganded Fab 2A2 crystals. This is different from the Fab 2A2-SpA complex crystallization, in which crystals grew under conditions that were nearly identical to those for the unbound Fab 2A2. Crystals for data collection were enlarged by macroseeding [37].

For data collection, crystals were transferred into a solution containing 50% reservoir solution and 50% cryo-solution (25%[wt/wt] MPEG 5,000, 25%(v/v) ethylene glycol, 9%(v/v) xylitol, 100 mM HEPES, pH 7.5). Data were recorded at 120°K from a single crystal on DW32 beamline at LURE synchrotron facility and processed using the HKL package [38]. The crystal belongs to the orthorhombic space group $P2_12_12_1$ with $a = 55.2$ Å, $b = 87.3$ Å and $c = 210.5$ Å. Statistics are summarized in Table 5.

### Structure Determination

We used the molecular replacement method to solve the crystal structure with the program AMoRe [39] by using coordinates of Fab 2A2 (PDB code: 1DEE) [13]. The variable ($V_L$-$V_H$) and the constant ($C_L$-$C_H1$) regions were used separately as search models for the corresponding parts of the Fab 2A2-PpL complex (Table 5).

After the initial refinement of this model, $\sigma_A$-weighted $2F_o - F_c$ and $F_o - F_c$ electron density maps were examined with the program O [40]. The difference maps show electron density for a single PpL domain sandwiched between two Fab fragments. Both the NMR structure of the B1 domain [7] and the crystal structure of the C* domain of PpL (B.J.S, unpublished) were fitted globally in the electron density maps with the best fit for the crystal structure. The 2 Fab:1 PpL model was refined within the 20–2.7 Å resolution range with the program CNS [41] and was rebuilt in O [40]. Detectable PpL domain C* residues range from Glu-20 to Ala-80 (see Figure 2 for numbering). As already observed by NMR [7], the 19 amino acid long N terminus is disordered and thus absent from the final structure. Small loops from the $C_H1$ region (H136 to H143 and H196 to H204) of both Fab molecules were poorly defined in maps and removed from the final structure. However, the $V_L$ and $V_H$ regions of both Fabs, the PpL domain, and in particular the $V_L$-PpL domain interfaces are well defined in the electron density maps (Figure 1b). Statistics for the structure determination are summarized in Table 5.

### Mutagenesis

The Y64W PpL and the Y53F-Y64W PpL mutants were constructed with the primers and methods described previously [21]. The D55A-Y64W mutant was made by the mutation of residue 55 on the Y64W template DNA by the use of the antisense primer 5′ TGC TAA TAA AGC TGC ATA TCT 3′. The site of the mutation is shown in bold. All mutations were confirmed by DNA sequencing with the Sequenase 2.2 DNA sequencing kit (United States Biochemical Corporation). Kd values determined at pH 8 were obtained by stopped-flow fluorescence measurements as previously described [14].

### Acknowledgments

We thank the Laboratoire pour l'Utilisation du Rayonnement Electromagnétique synchrotron facility (Orsay, France) for making beamline DW32 available to us as well as W. Shepard for his assistance and M. Hamon for cell culture. The 2A2 cell line was kindly provided by Professor R. Malpi (Kennedy Institute, London). We thank B. Gilquin and Professor A. Ménez for continuous support, M-H. LeDu, J. Menetrey, M. Knossow, and Professor J. Janin for reading the manuscript. E.A.S and B.J.S. acknowledge support from the Alliance Franco-British joint research program. B.J.S. and M.J.T. were supported by the Arthritis Research Campaign (United Kingdom) and by the Biotechnology and Biological Sciences Research Council (BBSRC; United Kingdom). Work at the Babraham Institute is supported by the BBSRC (United Kingdom). M.C.G. and B.J.S. thank the BBSRC and the Wellcome Trust for their support.

Received: March 22, 2001
Revised: June 1, 2001
Accepted: June 5, 2001

## References

1. Björck, L. (1988). Protein L. A novel bacterial cell wall protein with affinity for Ig L chains. J. Immunol. *140*, 1194–1197.
2. Romagnani, S., Giudizi, M.G., del Prete, G., Maggi, E., Biagiotti, R., Almerigogna, F., et al. (1982). Demonstration on protein A of two distinct immunoglobulin-binding sites and their role in the mitogenic activity of *Staphylococcus aureus* Cowan I on human B cells. J. Immunol. *129*, 596–602.
3. Björck, L., and Kronvall, G. (1984). Purification and some properties of streptococcal protein G, a novel IgG-binding reagent. J. Immunol. *133*, 969–974.
4. Kastern, W., Holst, E., Nielsen, E., Sjöbring, U., and Björck, L. (1990). Protein L, a bacterial immunoglobulin-binding protein and possible virulence determinant. Infect. Immun. *58*, 1217–1222.
5. Murphy, J.P., Duggleby, C.J., Atkinson, M.A., Trowern, A.R., Atkinson, T., and Goward, C.R. (1994). The functional units of a peptostreptococcal protein L. Mol. Microbiol. *12*, 911–920.
6. Kastern, W., Sjöbring, U., and Björck, L. (1992). Structure of peptostreptococcal protein L and identification of a repeated immunoglobulin light chain-binding domain. J. Biol. Chem. *267*, 12820–12825.
7. Wikström, M., Drakenberg, T., Forsén, S., Sjöbring, U., and Björck, L. (1994). Three-dimensional solution structure of an immunoglobulin light chain-binding domain of protein L. Comparison with the IgG-binding domains of protein G. Biochemistry *33*, 14011–14017.
8. Nilson, B.H., Solomon, A., Björck, L., and Åkerström, B. (1992). Protein L from *Peptostreptococcus magnus* binds to the kappa light chain variable domain. J. Biol. Chem. *267*, 2234–2239.
9. De Chateau, M., Nilson, B.H., Erntell, M., Myhre, E., Magnusson, G.G., Åkerström, B., et al. (1993). On the interaction between protein L and immunoglobuline of various mammalian species. Scand. J. Immunol. *37*, 399–405.
10. Patella, V., Casolaro, V., Björck, L., and Marone, G. (1990). Protein L. A bacterial Ig-binding protein that activates human basophile and mast cells. J. Immunol. *145*, 3054–3061.
11. Genovese, A., Bouvet, J.P., Florio, G., Lamparter-Schummert, B., Björck, L., and Marone, G. (2000). Bacterial immunoglobulin superantigen proteins A and L activate human heart mast cells by interacting with immunoglobulin E. Infect. Immun. *68*, 5517–5524.
12. Boder, E.T., Midelfort, K.S., and Wittrup, K.D. (2000). Directed evolution of antibody fragments with monovalent femtomolar antigen-binding affinity. Proc. Natl. Acad. Sci. USA *97*, 10701–10705.
13. Graille, M., Stura, E.A., Corper, A.L., Sutton, B.J., Taussig, M.J., Charbonnier, J.B., et al. (2000). Crystal structure of a *Staphylococcus aureus* protein A domain complexed with the Fab fragment of a human IgM antibody: structural basis for recognition of B-cell receptors and superantigen activity. Proc. Natl. Acad. Sci. USA *97*, 5399–5404.
14. Beckingham, J.A., Bottomley, S.P., Hinton, R., Sutton, B.J., and Gore, M.G. (1998). Interactions between a single immunoglobulin-binding domain of protein L from *Peptostreptococcus magnus* and a human kappa light chain. Biochem. J. *340*, 193–199.
15. Bottomley, S.P., Beckingham, J.A., Murphy, J.P., Atkinson, M., Atkinson, T., Hinton, R.J., et al. (1995). Cloning, expression and purification of Ppl-1, a kappa-chain binding protein, based upon protein L from *Peptostreptococcus magnus*. Bioseparation *5*, 359–367.
16. Kabat, E.A., and Wu, T.T. (1971). Attempts to locate complementarity-determining residues in the variable positions of light and heavy chains. Ann. N.Y. Acad. Sci. *190*, 382–393.
17. Al-Lazikani, B., Lesk, A.M., and Chothia, C. (1997). Standard conformations for the canonical structures of immunoglobulins. J. Mol. Biol. *273*, 927–948.
18. MacCallum, R.M., Martin, A.C., and Thornton, J.M. (1996). Antibody-antigen interactions: contact analysis and binding site topography. J. Mol. Biol. *262*, 732–745.
19. Åkerström, B., and Björck, L. (1989). Protein L: an immunoglobulin light chain-binding bacterial protein. Characterization of binding and physicochemical properties. J. Biol. Chem. *264*, 19740–19746.
20. Wikström, M., Sjöbring, U., Drakenberg, T., Forsén, S., and Björck, L. (1995). Mapping of the immunoglobulin light chain-binding site of protein L. J. Mol. Biol. *250*, 128–133.
21. Beckingham, J.A., Housden, N.G., Muir, N.M., Bottomley, S.P., and Gore, M.G. (2001). Studies on a single immunoglobulin-binding domain of protein L from *Peptostreptococcus magnus*: the role of tyrosine-33 in the reaction with human IgG. Biochem. J. *353*, 395–401.
22. Fersht, A.R. (1987). The hydrogen bond in molecular recognition. TIBS *12*, 301–304.
23. Conte, L.L., Chothia, C., and Janin, J. (1999). The atomic structure of protein-protein recognition sites. J. Mol. Biol. *285*, 2177–2198.
24. Janin, J. (1997). Specific versus non-specific contacts in protein crystals. Nat. Struct. Biol. *4*, 973–974.
25. de Vos, A.M., Ultsch, M., and Kossiakoff, A.A. (1992). Human growth hormone and extracellular domain of its receptor: crystal structure of the complex. Science *255*, 306–312.
26. Cunningham, B.C., Ultsch, M., De Vos, A.M., Mulkerrin, M.G., Clauser, K.R., and Wells, J.A. (1991). Dimerization of the extracellular domain of the human growth hormone receptor by a single hormone molecule. Science *254*, 821–825.
27. Harris, L.J., Larson, S.B., Hasel, K.W., and McPherson, A. (1997). Refined structure of an intact IgG2a monoclonal antibody. Biochemistry *36*, 1581–1597.
28. Harris, L.J., Skaletsky, E., and McPherson, A. (1998). Crystallographic structure of an intact IgG1 monoclonal antibody. J. Mol. Biol. *275*, 861–872.
29. Goward, C.R., Scawen, M.D., Murphy, J.P., and Atkinson, T. (1993). Molecular evolution of bacterial cell-surface proteins. Trends Biochem. Sci. *18*, 136–140.
30. Derrick, J.P., and Wigley, D.B. (1994). The third IgG-binding domain from streptococcal protein G. An analysis by X-ray crystallography of the structure alone and in a complex with Fab. J. Mol. Biol. *243*, 906–918.
31. Derrick, J.P., and Wigley, D.B. (1992). Crystal structure of a streptococcal protein G domain bound to an Fab fragment. Nature *359*, 752–754.
32. Derrick, J.P., Feavers, I., and Maiden, M.C. (1999). Use of streptococcal protein G in obtaining crystals of an antibody Fab fragment in complex with a meningococcal antigen. Acta Crystallogr. D Biol. Crystallogr. *55*, 314–316.
33. Stura, E.A., Graille, M., and Charbonnier, J.-B. Crystallization of macromolecular complexes: combinatorial complex crystallization. J. Cryst. Growth, in press.
34. Ostermeier, C., Iwata, S., Ludwig, B., and Michel, H. (1995). Fv fragment-mediated crystallization of the membrane protein bacterial cytochrome c oxidase. Nat. Struct. Biol. *2*, 842–846.
35. Sohi, M.K., Sutton, B.J., Corper, A.L., Wan, T., Meini, R.N., Brown, C., et al. (1994). Crystallization and preliminary X-ray analysis of the Fab fragment of a human monoclonal IgM rheumatoid factor (OAZ). J. Mol. Biol. *242*, 706–708.
36. Stura, E.A., Graille, M., Taussig, M.J., Sutton, B.J., Gore, M.G., and Charbonnier, J.B. Crystallization of macromolecular complexes: stoichiometric variation screening. J. Cryst. Growth., in press.
37. Stura, E.A., and Wilson, I.A. (1991). Seeding techniques. 241–254.
38. Otwinowski, Z., and Minor, W. (1997). Methods Enzymol. *276*, 307–326.
39. Navaza, J. (1994). AMoRe: an Automated Package for Molecular Replacement. Acta Crystallogr. A *50*, 157–163.
40. Jones, T.A., Zou, J.Y., Cowan, S.W., and Kjeldgaard (1991). Improved methods for binding protein models in electron density maps and the location of errors in these models. Acta Crystallogr. A *47*, 110–119.
41. Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleva, R.W., et al. (1998). Crystallography + NMR

system: A new software suite for macromolecular structure determination. Acta Crystallogr. D *54*, 905–921.
42. Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph *14*, 51–5, 29–32.
43. Roussel, A., and Cambillau, C. (1989). Silicon Graphics Geometry Partner Directory. (Mountain View, CA: Silicon Graphics), pp. 77–79.
44. Bailey, S. (1994). The CCP4 Suite: Programs for protein crystallography. Acta Crystallogr. D *50*, 760–763.
45. McDonald, I.K., and Thornton, J.M. (1994). Satisfying hydrogen bonding potential in proteins. J. Mol. Biol. *238*, 777–793.

## Accession Numbers

Coordinates and structure factors have been deposited in the Protein Data Bank (accession code: 1HEZ).

## Note Added in Proof

After we submitted this paper, Ulf Sjöbring informed us that his laboratory conducted extensive site-directed mutagenesis to document PpL domain folding in collaboration with D. Baker. These experiments, in good agreement with our X-ray structure, gave strong indications that there are two binding sites in protein L: one in the β2-alpha helix and another one in the β3 alpha helix (facing the other way).

# Partially Folded Intermediates as Critical Precursors of Light Chain Amyloid Fibrils and Amorphous Aggregates†

Ritu Khurana, Joel R. Gillespie,‡ Anupam Talapatra, Lauren J. Minert, Cristian Ionescu-Zanetti, Ian Millett,‡ and Anthony L. Fink*

*Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064*

*Received July 31, 2000; Revised Manuscript Received January 22, 2001*

ABSTRACT: Light chain, or AL, amyloidosis is a pathological condition arising from systemic extracellular deposition of monoclonal immunoglobulin light chain variable domains in the form of insoluble amyloid fibrils, especially in the kidneys. Substantial evidence suggests that amyloid fibril formation from native proteins occurs via a conformational change leading to a partially folded intermediate conformation, whose subsequent association is a key step in fibrillation. In the present investigation, we have examined the properties of a recombinant amyloidogenic light chain variable domain, $SMA_v$, to determine whether partially folded intermediates can be detected and correlated with aggregation. The results from spectroscopic and hydrodynamic measurements, including far- and near-UV circular dichroism, FTIR, NMR, and intrinsic tryptophan fluoreseence and small-angle X-ray scattering, reveal the build-up of two partially folded intermediate conformational states as the pH is decreased (low pH destabilized the protein and accelerated the kinetics of aggregation). A relatively nativelike intermediate, $I_N$, was observed between pH 4 and 6, with little loss of secondary structure, but with significant tertiary structure changes and enhanced ANS binding, indicating exposed hydrophobic surfaces. At pH below 3, we observed a relatively unfolded, but compact, intermediate, $I_U$, which was characterized by decreased tertiary and secondary structure. The $I_U$ internediate readily forms amyloid fibrils, whereas $I_N$ preferentially leads to amorphous aggregates. Except at pH 2, where negligible amorphous aggregate is formed, the amorphous aggregates formed significantly more rapidly than the fibrils. This is the first indication that different partially folded intermediates may be responsible for different aggregation pathways (amorphous and fibrillar). The data support the hypothesis that amyloid fibril formation involves the ordered self-assembly of partially folded species that are critical soluble precursors of fibrils.

Immunoglobulin (Ig)[1] light chains are involved in several protein deposition diseases, including one resulting in the formation and deposition of amyloid fibrils (light chain or AL amyloidosis) and another known as light chain deposition disease that involves amorphous protein deposits (*1*, *2*). The morphology of the deposited aggregates in these two diseases is clearly different, and typically patients exhibit only one form of light chain deposition. However, there is at least one report of a patient exhibiting both AL amyloidosis and LCDD involving the same light chain (*3*). The exact length of light chains in amyloid deposits varies, but is usually in the 110−130 residue range (12−14 kDa) corresponding to the intact variable domain (*4*).

The molecular mechanisms leading to amyloid formation are poorly understood. In this report, we address the question of why some immunoglobulin light chains form amyloid and related deposits while others do not, in particular the hypothesis that protein aggregation arises from the self-association of partially folded intermediates. Support for this hypothesis has been found with proteins such as transthyretin (*5*) and lysozyme (*6*). We postulate that amyloid fibril formation from native proteins occurs via a conformational change leading to formation of a partially folded intermediate conformation, association of this internediate to form soluble oligomers leading to the critical nucleus, and subsequent formation of the initial fibrillar species, typically a filament or protofibril, and finally association of protofibrils into mature fibrils (*7*).

We have investigated the biophysical properties and amyloidogenicity of the variable domain of a recombinant amyloidogenic light chain, SMA, engineered by Stevens et al. (*8*). SMA (114 residues, $M_r$ = 12 700) was initially

† This research was supported by grants from the National Institutes of Health and the University of California Systemwide Biotechnology Research and Education Program. Small-angle X-ray scattering data were collected at Beam Line 4-2, Stanford Synchrotron Radiation Laboratory (SSRL). SSRL is supported by the U.S. Department of Energy, Office of Basic Energy Sciences, and in part by the National Institutes of Health, National Center for Research Resources, Biomedical Technology Program, and by the Department of Energy, Office of Biological and Environmental Research.

* Correspondence should be addressed to this author at the Department of Chemistry and Biochemistry, University of California, Santa Cruz, CA 95064. Tel: (831) 459-2744; Fax: (831) 459-2935; E-mail: cnzyme@cats.ucsc.edu.

‡ Current address: Department of Chemistry, Stanford University, Stanford, CA 94305.

[1] Abbreviations: SAXS, small-angle X-ray scattering; $R_g$, radius of gyration; FTIR, Fourier transform infrared spectroscopy; ATR, attenuated total reflectance; IRE, internal reflectance element; TEM, transmission electron microscopy; AFM, atomic force microscopy; TFT, Thioflavin T; ANS, 8-anilino-1-naphthalenesulfonate; CD, circular dichroism; Ig, immunoglobulin; $V_L$, variable domain of Ig light chain.

10.1021/bi001782b CCC: $20.00 
Published on Web 03/02/2001

extracted from lymph node-derived amyloid fibrils of an AL amyloidosis patient (*9*). A very similar light chain domain, LEN, was derived from a patient with multiple myeloma who showed no evidence of renal dysfunction or amyloidosis (*10*). The IgG-$V_L$ domains consist of a highly conserved framework formed by two sheets of antiparallel $\beta$-strands forming a $\beta$-sandwich, and three loops comprising the complementarity-determining regions (CDR) that form part of the antigen binding site. The sequences of SMA and LEN are very similar, differing only at 8 positions out of 114. Four of these are in CDR3 (Q89H, T94H, Y96Q, S97T), two are in CDR1 (S29N, K30R), and the remaining two are in the framework region (P40L, I106L). The high-resolution crystallographic structure of LEN (1.8 Å) has been solved (PDB Accession No. 1LVE) (*11*). Both SMA and LEN belong to the $\kappa$IV family of Igs.

The amyloidogenic light chain, SMA, is significantly less thermodynamically stable than LEN under all conditions (unpublished observations). The presence of low concentrations of denaturants also results in fibril formation from the "benign" LEN (*12*). In the present study, we used biophysical characterization of the conformation of SMA as a function of pH to reveal the presence of two distinct partially folded intermediates: one with relatively nativelike properties, the other with relatively unfolded properties. Destabilizing conditions at physiological pH, e.g., low urea concentration, also lead to aggregation and fibril formation. Thus, conditions that result in population of these intermediates lead to aggregation, supporting the hypothesis that partially folded intermediates are key precursors on the aggregation pathway. Interestingly, both amorphous and fibrillar aggregates were observed, and were shown to arise from two different intermediates.

## MATERIALS AND METHODS

*Expression and Purification of Recombinant $V_L$ SMA.* The recombinant $V_L$ domain SMA was purified from JM83 *E. coli* cells transformed with the plasmid pkIVsma004, generously provided by Dr. Fred Stevens, Argonne National Lab (*8*). The plasmid construct was based on the pASK vectors constructed by Skerra et al. (*13*) and contained an *ompA* leader for periplasmic localization of the protein to ensure the formation of the core disulfide bond. The overexpressed protein was purified using the procedure of Stevens et al. (*8*) with minor modifications. Briefly, the recombinant protein was extracted from the periplasm using osmotic shock via treatment with ice-cold TES followed by distilled water. The periplasmic extract was dialyzed against 4 changes of 20 volumes of 10 mM acetate buffer, pH 5.6, and loaded onto a fast-flow SP Sepharose column. The column was washed with 10 mM acetate buffer, pH 5.6, and the protein eluted using 10 mM phosphate buffer, pH 8.0. The fractions were assayed by SDS−PAGE, and fractions containing the recombinant protein were pooled, filtered through 0.22 $\mu$m filters, and stored in glass vials. Typical yields were 7−8 mg of purified protein per liter of cells. Protein concentrations were measured via optical density at 280 nm using the extinction coefficient of $E_{0.1\%} = 1.8$ calculated from the sequence. The purified protein was stored in 10 mM phosphate buffer (pH 8.0) at 4 °C and used within 2 weeks of the initial purification. The purity of the protein preparations was assayed by SDS−PAGE and by electrospray mass spectrometry (Micromass Quattro II).

*Intrinsic Tryptophan Fluorescence Measurements.* Fluorescence measurements were made with a FluoroMax-2 fluorescence spectrometer (Jobin Yvon-Spex). Emission spectra between 300 and 420 nm were collected with excitation at 280 nm. Spectra were collected at different pHs within the range of 10−2 using 0.5 $\mu$M protein samples in 50 mM of the appropriate buffer containing 100 mM NaCl. Spectra were collected at 25 and 37 °C. The stability of SMA toward urea denaturation was monitored as a function of pH by recording changes in tryptophan fluorescence intensity upon excitation at 280 nm and emission at 350 nm at 25 °C.

Samples of SMA (1 $\mu$M monomer) were incubated in 20 mM phosphate buffer (pH 7.4), 100 mM NaCl containing varying amounts of urea (0−8 M) for 2 h to ensure completion of the unfolding reaction. Data were analyzed by nonlinear least-squares fitting to a two-state folding model. The fraction unfolded, $F_u$, was calculated using the equation: $F_u = (y_f - y)/(y_f - y_u)$ where $y$ represents the observed fluorescence at a particular concentration of urea, and $y_f$ and $y_u$ represent the corresponding fluorescence of the folded and unfolded states, respectively, at that urea concentration. For baseline fitting, a linear least-squares analysis was performed to determine the values of $y_f$ and $y_u$ as a function of urea concentration. The free energies of unfolding were calculated as a function of urea concentration using the equation: $\Delta G = -RT \ln K_{eq}$, where $K_{eq} = f_u/(1 - f_u)$. $\Delta G°$ was determined by linear extrapolation to zero urea concentration using the expression: $\Delta G° = \Delta G + m[\text{urea}]$.

*ANS Binding.* 1,8-Anilinonaphthalenesulfonate was obtained from Kodak, and a stock solution was prepared by dissolving in water followed by filtration through a 0.2 $\mu$m syringe filter and measuring the concentration using an extinction coefficient of 5000 $M^{-1}$ $cm^{-1}$ at 350 nm. The fluorescence emission spectra of solutions of 10 $\mu$M ANS and 0.5 $\mu$M protein were collected between 420 and 600 nm upon excitation at 380 nm as a function of pH at 37 °C.

*Circular Dichroism Spectra.* CD spectra were collected on an AVIV model 62DS spectrometer between 260 and 190 nm for the far-UV region and between 320 and 250 nm for the near-UV region, with a step size of 0.5 nm and an averaging time of 5 s and collecting 5 repeat scans. Cells of 1 and 0.01 cm path length were used for near- and far-UV CD measurements with protein concentrations of 0.5 and 1.7 mg/mL, respectively.

*pH Dependence.* pH-dependent changes in spectroscopic data were fit using a modified Henderson−Hasselbalch equation for one (eq 1) or two (eq 2) transitions, to determine the midpoints of the transitions:

$$Y_{obs} = \frac{Y_N + Y_{I_2} 10^{pH - pH_{m2}}}{1 + 10^{pH - pH_{m2}}} \quad (1)$$

$$Y_{obs} = \frac{Y_N + Y_{I_N} 10^{pH - pH_{m1}} + \dfrac{Y_{I_C}}{10^{pH - pH_{m2}}}}{1 + 10^{pH - pH_{m2}} + \dfrac{1}{10^{pH - pH_{m2}}}} \quad (2)$$

where $Y_{obs}$ is the observed spectroscopic property. $Y_N$ is the

value of the spectroscopic property for the native state. $Y_{I_N}$ is the spectroscopic property for the nativelike intermediate, and $Y_{I_U}$ is the spectroscopic property for the unfolded-like intermediate. $pH_{m1}$ and $pH_{m2}$ are the midpoints of transitions from the native state to the $I_N$ intermediate and from $I_N$ to the $I_U$ intermediate, respectively (*36*).

*Thin film ATR-FTIR measurements* were performed using a SPECAC out-of-compartment ATR accessory and a Nicolet 800SX FTIR bench. A germanium crystal IRE was used for making hydrated thin films of ~50−100 $\mu$g of protein from both soluble and insoluble protein as previously described (*14, 15*). ATR-FTIR spectra were collected followed by Fourier transformation of the sample spectra using a clean crystal spectrum as a background. The water vapor spectrum was collected by reducing the air purge and subtracted from the protein spectrum until the spectra were featureless in the region between 1700 and 1800 $cm^{-1}$. ATR-FTIR spectra for SMA were collected at pH 7.5 in 50 mM sodium phosphate, 100 mM NaCl, at pH 5.0 in 50 mM sodium acetate, 100 mM NaCl, and at pH 2.0 in 20 mM HCl, 100 mM NaCl. Buffer spectra were subtracted from the sample spectra. Component spectra were obtained by first determining peak positions using both second-derivative and Fourier-self-deconvoluted spectra, followed by curve-fitting to the raw spectrum (*16*).

*In Vitro Fibril Formation Assays.* Fibril formation was monitored using a fluorescence assay based on the enhanced fluorescence of the dye Thioflavin T (TFT) on binding to amyloid fibrils (*17*). Amyloid fibrils were grown from purified protein (40 $\mu$M) in 50 mM buffer and 100 mM NaCl. A filtered protein sample (using 0.22 $\mu$m syringe filters) was incubated under the desired conditions in a 1.8 mL flat-bottomed screw-capped glass vial with moderate stirring using a Teflon-coated micro-stir bar. Typical fibril growth experiments involved incubating the protein at 37 °C with constant stirring and removing aliquots (10 $\mu$L) over time for analysis by light scattering and TEM and TFT binding (see below). Standard buffers included 20 mM HCl or phosphate (pH 2), 50 mM formate (pH 3 and 4), 50 mM cacodylate or acetate (pH 5 and 6), and 20 mM TRIS or 20 mM HEPES or 50 mM phosphate (pH 7). Both Rayleigh light scattering and fluorescence spectra were collected using a SPEX/Jobin-Yvon Fluoromax-2 spectrofluorometer. Constant temperatures were maintained using a circulating water bath. At each time point, a 220 $\mu$L sample was removed and transferred to a cylindrical quartz microcell to measure Rayleigh light scattering at 330 nm with a 1 nm band-pass for both excitation and emission monochromators. Thioflavin T binding assays were conducted by adding sample aliquots (10 $\mu$L) to 990 $\mu$L of 20 $\mu$M TFT in 50 mM TRIS, pH 7.5, and 100 mM NaCl in a 1 mL fluorescence cuvette. Fluorescence emission was monitored with excitation at 450 nm using a 5 nm band-pass on both the excitation and emission monochromators. Fluorescence intensities were reported at 482 nm.

*SAXS Measurements.* Small-angle X-ray scattering measurements were performed on beam line 4-2 at the Stanford Synchrotron Radiation Laboratory (SSRL) as described previously (*18*). The SAXS instrument was configured with a Mo:$CB_4$ multilayer monochromator, an 18 mm beamstop, and a 218 cm sample-to-detector distance. Data were collected on protein samples ranging from 0.5 to 10 mg/mL in 50 mM buffer and 100 mM NaCl at 37 °C using a PTFE flow-cell with 1.3 mm path length to minimize radiation damage. Radii of gyration were calculated using the Guinier approximation (*19*).

*Atomic Force and Transmission Electron Microscopy.* Transmission electron micrographs were collected using a JEOL JEM-100B microscope operating with an accelerator voltage of 80 kV. Typical nominal magnifications ranged from 27000× to 67000×. Samples were deposited on Formvar-coated 300 mesh copper grids and negatively stained with freshly prepared 2% aqueous uranyl acetate.

For AFM, aliquots of 50 $\mu$L of incubation solution were transferred to an Eppendorf tube and spun to pellet the precipitated material, which was then washed twice with water before resuspending in deionized water. A drop of aggregate suspension was deposited on freshly cleaved mica and dried immediately with nitrogen gas. The samples were imaged with an Autoprobe CP AFM (Park Scientific, Sunnyvale, CA) in the noncontact (NC-AFM) mode. The tube scanner was a 100 $\mu$m Scanmaster (Park Scientific). NC Ultralevers (Park Scientific) were used as cantilevers. The resonant frequency was ~100 kHz. The images were taken in air, ambient conditions, at a scan frequency of 1−2 Hz, using silicon nitride tips.

*NMR Spectroscopy.* $^1H$ NMR spectra were collected using a Varian Unity+ 500 spectrometer equipped with ultrashims and a Varian triple resonance probe. Presaturation and postacquisition digital filtering were used for solvent suppression. Data were collected on 0.5 mM protein samples containing 100 mM NaCl and 10% $D_2O$. The pH was adjusted by titration with NaOH or HCl as needed. All data were recorded at 37 °C.

*Light Scattering Measurements at High SMA Concentration.* The rate of aggregation was monitored by static light scattering using a Nepheloskan instrument from Labsystems and a 96-well plate reader. Solutions of 3.5 and 7 mg/mL SMA at the appropriate pH were incubated at 37 °C along with their corresponding buffers, and scattering was measured every 15 min for 6 h.

*pH Jump Experiments.* Interconversions between N, $I_N$, $I_D$, and U were monitored by diluting 10 $\mu$M SMA at one pH into buffer of another pH, such that the final concentration of protein was 1 $\mu$M. After manually mixing, the intensity of either tryptophan fluorescence (excitation 280 nm and emission 345 nm) or ANS fluorescence (excitation 380 nm and emission 470 nm) was monitored, using a time-based scan on a Spex Fluoromax instrument with 1 s time-averaging.

## RESULTS

*Formation of Amyloid Fibrils Is Favored by Destabilizing Environmental Conditions.* The enhanced fluorescence emission of the dye Thioflavin T on association with amyloid fibrils provides a very convenient method to monitor the kinetics of amyloid fibril formation (*17, 20*). Fibril formation by SMA was investigated by stirring solutions of SMA at various values of pH at 37 °C. At room temperature, or in the absence of stirring, no enhancement of TFT was observed for several days.

The rate of fibril formation from SMA was found to be very sensitive to a number of extrinsic factors, including pH,

agitation, and temperature. As is typical for other amyloid fibril growth curves, SMA fibrillation kinetics exhibit a quasi-sigmoidal behavior, consisting of a lag phase followed by a logarithmic growth phase that eventually plateaus (Figure 1). The slight drop in TFT fluorescence sometimes observed at long time periods may reflect conversion of the mature SMA amyloid fibrils to an alternative fibrillar form that has a lower affinity toward TFT. For example, we frequently observed lateral aggregation of mature fibrils coinciding with the decrease in TFT fluorescence, suggesting a possible decrease in the availability of TFT binding sites.

We confirmed that the enhanced fluorescence emission of TFT was indeed due to interaction with SMA amyloid fibrils by the characteristic Congo red green-birefringence observed under crossed-polarization (data not shown) and by direct observation with transmission electron microscopy and atomic force microscopy. These techniques demonstrated the presence of fibrils in systems with high TFT fluorescence, and their absence in samples with no increase in TFT fluorescence. The TFT assay was shown to be linear over the concentration range of 0 to >12 $\mu$g of amyloidogenic protein in the assay solution.

The initial lag in fibril formation (see Figure 1) is often attributed to the slow assembly of a critical nucleus in a nucleation−polymerization mechanism (*21*). The length of the lag (measured by extrapolating the exponential growth phase to zero intensity) during SMA fibril formation was a sensitive function of the pH at which the fibrils were generated. Some typical data are shown in Figure 1. The length of the lag decreased from days at pH 7.0, 37 °C ,for 40 $\mu$M SMA, to a few hours at pH 2 (these values for the lag time are very sensitive to the rate of stirring or agitation). In addition, the maximum signal obtained using the TFT assay increased with decreasing pH, indicating that more fibrils were formed at lower pH values (note that the TFT binding assay is performed at pH 7).

Static light scattering was also used to monitor the kinetics of aggregation (Figure 1). Surprisingly, we noted that amorphous aggregation occurred in the same incubation samples of SMA as fibril formation, but at a faster rate. Substantial amorphous aggregation was observed from pH 7 to 4. The amount of amorphous aggregation was proportional to the protein concentration. The amorphous aggregation was observed immediately after starting the stirring at 37 °C as indicated by increased light scattering (Figure 1B,C), whereas the presence of fibrils, as reflected by the increase in TFT fluorescence, was not observed for several days. Confirmation of the fact that this early aggregation was indeed amorphous comes from TEM and AFM micrographs that showed amorphous material and the absence of fibrils (Figure 2). Under certain conditions, e.g., pH 5.0, 37 °C, 100 mM NaCl, 50−60% of the SMA had precipitated (as amorphous aggregate) after 24 h of incubation (based on the absorbance of the supernatant), and no fibrils were visible by microscopy. In contrast, at pH ≤3 essentially all of the precipitate was fibrillar within 24 h. The rate of the increase in the light scattering observed at pH 2 correlated with that of the increase in Thioflavin T fluorescence (Figure 1A), suggesting that the predominant species present in solution was fibrillar. This was confirmed by TEM, which indicated that the aggregates were largely fibrillar, though some amorphous material was present. Interestingly the maximum increase in Thioflavin T fluorescence was significantly less at pH 7 compared to pH 5 and pH 2, which also correlated with fewer fibrils observed by microscopy at pH 7 compared to the lower pH conditions.

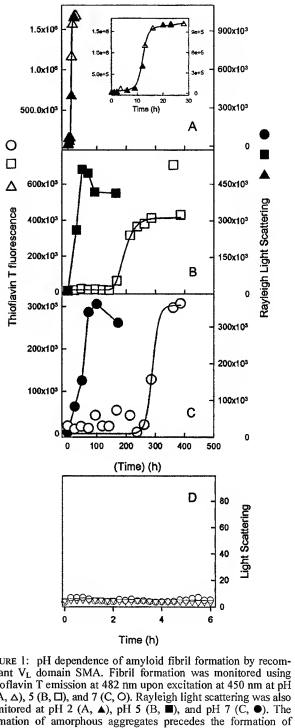



FIGURE 1: pH dependence of amyloid fibril formation by recombinant $V_L$ domain SMA. Fibril formation was monitored using Thioflavin T emission at 482 nm upon excitation at 450 nm at pH 2 (A, △), 5 (B, □), and 7 (C, ○). Rayleigh light scattering was also monitored at pH 2 (A, ▲), pH 5 (B, ■), and pH 7 (C, ●). The formation of amorphous aggregates precedes the formation of amyloid fibrils at pH 7 and 5, conditions that favor the native or the nativelike intermediate conformation. The inset to panel A shows an expanded time scale. Light scattering is sensitive to the presence of both amorphous and fibrillar material, whereas TFT fluorescence is selective for fibrillar aggregates alone. Panel D shows that in the absence of agitation, at SMA concentrations as high as 0.5 mM, no aggregation occurs over at least a 6 h period: circles are for pH 7, triangles for pH 2. With agitation, the signal would be >1400 due to the aggregation.

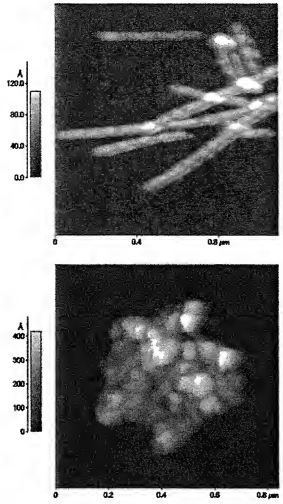



FIGURE 2: Atomic force microscopy images of SMA fibrils observed at pH 2 after 24 h of stirring at 37 °C (A, top). At pH 5 and 7 after incubation for 1 h at 37 °C with constant stirring, a nebulous and loosely packed amorphous deposit is observed (B, bottom). After a few days at pH 5 and 7, both fibrils and amorphous deposits are observed.

*Morphology of Fibrillar and Amorphous Deposits.* The SMA aggregates were examined using atomic force (Figure 2) and transmission electron microscopy. The images revealed unbranched, rope-like fibrils (Figure 2A), several hundred nanometers in length, most with diameters of ~8 nm, but some, protofibrils, with diameters of ~4 nm (*7*). Upon incubation at 37 °C at pH 4−6, amorphous aggregates were observed (Figure 2B).

*Spectroscopic Characterization of Acidic pH Conformations of SMA.* A number of spectroscopic probes were utilized to examine conformational changes in SMA that occur under conditions favoring the formation of amorphous aggregates (pH 4−6) and amyloid fibrils (pH <3). Both amyloid fibrils and amorphous aggregates are only formed in solutions of SMA at 37 °C upon agitation of the solution. Note that all the spectroscopic analyses were done at 37 °C without agitation within 2−3 h of preparation, ensuring that only soluble equilibrium conformations were studied; i.e., neither amorphous nor fibrillar aggregates were present in the spectroscopic analyses. No light scattering was observed for at least 6 h for solutions of SMA at concentrations as high as 7 mg/mL (used in the NMR experiments) at various pH

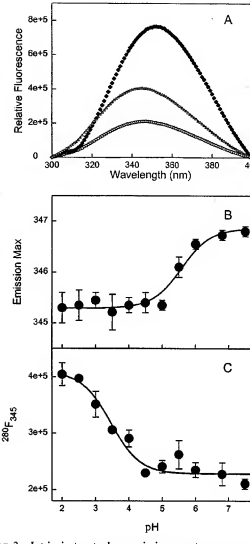



FIGURE 3: Intrinsic tryptophan emission spectra were measured with excitation at 280 nm for 0.5 μM protein solution in 50 mM buffer and 100 mM NaCl at 37 °C for native SMA at pH 7.5 (◆), denatured SMA in 8 M urea at pH 7.5 (○), and $I_U$ at pH 2 (△). The wavelength of maximum emission of tryptophan fluorescence (panel B) and emission intensity (panel C) are plotted against pH. The solid lines are fits to a single ionizable group pH transition using eq 1 (see Materials and Methods). The midpoint of pH transitions for emission maximum and the intensity at 345 nm were 5.6 and 3.5, respectively.

values from 2 to 7 (Figure 1D). Tertiary structure changes were monitored by tryptophan fluorescence emission, near-UV CD, far-UV CD (via the 230 nm peak resulting from aromatic clustering), and by ANS binding studies. Secondary structure changes were monitored by far-UV CD and Fourier transform infrared spectroscopy.

The two tryptophan residues of SMA, W35 and W50, provided convenient spectroscopic means for assessing the protein's conformational state. In particular, W35, which was quenched by the close proximity of the core disulfide formed from C23 and C88 in the native state, was observed to provide a probe of the global conformational state of the $V_L$ domain. Unfolding of the protein resulted in a decrease of the quenching and a consequent increase in the emission intensity of W35 (Figure 3A, and ref *31*). The second tryptophan residue, W50, was solvent-exposed in the native

state, based on both the crystal structure of LEN and the observed $\lambda_{max}$ of 347 nm in native SMA and LEN.

The intrinsic tryptophan fluorescence spectra indicated $\lambda_{max}$ of 347 and 355 nm for native and denatured SMA (8 M urea), respectively (Figure 3A). In addition, denatured SMA showed a large increase in Trp fluorescence intensity compared to that of its native conformation. When the pH of a solution of SMA was lowered from 7.5, significant changes were observed in the tryptophan fluorescence emission properties. In the pH 4−6 region, there was a decrease in $\lambda_{max}$, from 347 to 345 nm (Figure 3B), which was attributed to the build-up of a partially folded intermediate. For reasons to be discussed below, this intermediate was called $I_N$ (for nativelike intermediate). These data were fit to eq 1, and the midpoint of this pH transition was calculated to be 5.6. At pH <4, the emission intensity increases but the $\lambda_{max}$ does not change further (Figure 3B,C). The increase in emission intensity observed from pH 5 to 2 suggests significant disassembly of the hydrophobic core of the protein. The midpoint of this transition was pH 3.4 and appeared to be cooperative (Figure 3C). The fluorescence spectrum of SMA at pH 2 had a fluorescence intensity between those of the native and denatured states, and a blue shift in the maximum emission to 345 nm. This suggested that substantial residual structure was present at pH 2. This conformation of SMA, populated below pH 3, was called $I_U$. In addition, the fluorescence spectrum of SMA was measured as a function of protein concentration, over the 0.05−0.5 mg/mL range, to confirm the presence of the two intermediates and to demonstrate the lack of association of the samples under the experimental conditions.

The near-UV CD spectrum of SMA contained significant contributions from the aromatic (tryptophan, tyrosine, and phenylalanine) residues that were sensitive to the tertiary structure of the protein. The near-UV CD spectrum of native SMA (Figure 4A) showed positive peaks at 286 and 296 nm. These likely reflected contributions from two aromatic clusters involving residues Y36, Y86, Y87, F98 and Y27(d), Y32, Y49, Y91, Y92, observed in the crystal structure of LEN which would also be expected to be present in SMA. The peaks at 286 and 296 nm disappeared with transition midpoints of pH 3.2 and 3.4, respectively (Figure 4B). These transitions correspond to the formation of the $I_U$ species, suggesting loss of the nativelike environment of the aromatic groups in this intermediate. The small positive ellipticity for SMA at 268 nm showed transitions with midpoints of 4.9 and 3.7, corresponding to the transitions to $I_N$ and $I_U$, respectively (Figure 4B). At pH 2, the near-UV CD spectrum of SMA was essentially featureless (Figure 4A), suggesting loss of most of the tertiary structure in $I_U$, including the aromatic clusters. As a whole, the near-UV CD spectra for SMA in the pH 4−6 region indicated that the underlying tertiary structure is still relatively nativelike in this pH range, consistent with the presence of a nativelike conformation in $I_N$.

The hydrophobic dye ANS has frequently been used as a probe to reveal the presence of partially folded intermediates due to the presence of increased exposure of contiguous hydrophobic surface area (*22*−*24*). ANS did not significantly bind to SMA in its native state, indicating the absence of exposed hydrophobic pockets. However, a pH-titration of SMA in the presence of ANS at 37 °C revealed a marked

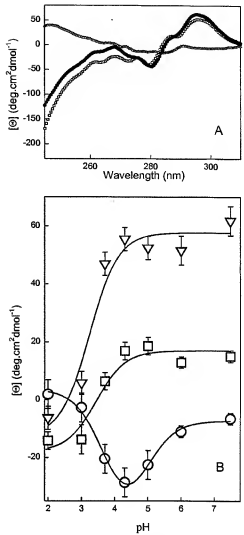



FIGURE 4: Near-UV CD spectra for SMA. Panel A shows plots of molar ellipticity for pH 7.5 (●), pH 5 (□), and pH 2 (△). Panel B shows the pH dependence of the molar ellipticity at 296 nm (▽), 287 nm (○), and 268 nm (□). The lines through the data are fits to eq 1 for peaks at 296 and 287 nm, yielding apparent $pK_a$s of 3.2 and 3.4, respectively, which correspond to formation of $I_U$. The data at 268 nm are fitted to eq 2, resulting in two apparent $pK_a$ values of pH 4.9 and 3.7. Protein concentration was 0.7 mg/mL.

increase in the fluorescence emission and a blue shift in the ANS emission maximum from 510 to 480 nm. Both the emission intensity increase (data not shown) and the blue shift of the emission were indicative of exposed hydrophobic regions in the vicinity of pH 4−6, with a maximum at pH 4.5 (Figure 5). Such an observation was taken to indicate the build-up of a partially folded intermediate, $I_N$. The midpoints of the transitions observed for SMA were at pH 5.2 and 3.8 (Figure 5). The limited ANS binding at low pH was attributed to the second intermediate, $I_U$, which appears to have less contiguous exposed hydrophobic regions. The data suggest that the $I_N$ intermediate is maximally populated between pH 4 and 5 for SMA. In contrast, no ANS binding was observed in the pH 2−10 range for the nonamyloidogenic homologue LEN, indicating the absence of both intermediates with LEN (Figure 5). In addition, the structure of LEN at pH 2 is nativelike, based on small-angle X-ray

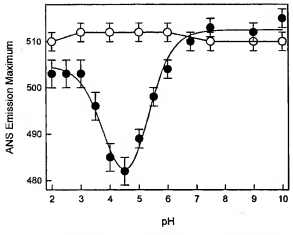



FIGURE 5: pH dependence of binding of ANS to SMA (●) and LEN (○). The reaction was monitored with 0.5 μM protein solution and 10 μM ANS, with excitation at 380 nm. Fluorescence emission spectra were collected between 400 and 600 nm at different pHs. The solid line is a fit to two transitions using eq 2 (see Materials and Methods). The midpoints of the pH transitions were 5.2 and 3.8, respectively.

scattering, circular dichroism, FTIR, and Trp fluorescence (unpublished observations).

Circular dichroism spectra were collected for SMA from pH 8 to 2 to probe global secondary structural changes in the different conformers. The far-UV circular dichroism spectrum of native SMA at pH 7.5 was rather unusual, in that it had distinct minima at 230 nm, as well as at 216 nm (Figure 6A). The former was attributed to contributions from aromatic interactions and possibly the disulfide, the latter to β-structure. When the CD spectrum of SMA was examined as a function of pH at 37 °C, there were relatively small changes between pH 7.5 and pH 4, with more significant changes occurring at lower pH (Figure 6). The former are consistent with loss of some tertiary structure in $I_N$. The spectrum at pH 2 was significantly different from the spectrum of SMA denatured in 7 M urea or 5 M Gdn·HCl at pH 7.4, indicating significant structure at pH 2. Plots of the ellipticity at 230 nm against pH reveal the population of a distinct conformational species in the pH 4−6 region (Figure 6B). The ellipticity of SMA monitored at 216 nm (corresponding to β-sheet structure) indicated no change between pH 7.5 and 4.0, suggesting that $I_N$ is relatively nativelike. However, below pH 4.0, the negative ellipticity at 216 nm shifted toward lower wavelengths, consistent with the loss of β-sheet structure (Figure 6C). The spectrum at pH 2.0 was a mixture of contributions from β-sheet and random coil conformations, indicating some loss of nativelike β-structure at pH 2. The data were consistent with the presence of a relatively unfolded-like intermediate, $I_U$, at pH below 3.

Fourier transform infrared spectroscopy (FTIR) has been used to probe protein structure, and the amide I band (1600−1700 $cm^{-1}$) has been used to estimate protein secondary structure content (*25*). The ATR-FTIR spectra of hydrated thin films of SMA at pH 7.5 and 5 revealed that significant secondary structural changes occur for the $I_N$ intermediate compared to the native state (Figure 7). The major differences are an increase in low-frequency β-sheet (1625 $cm^{-1}$), an increase in disordered structure (1648 $cm^{-1}$), increased turn

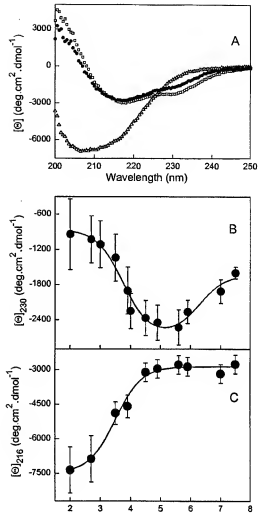



FIGURE 6: Far-UV CD spectra of SMA as a function of pH. Panel A shows the spectra at pH 7.5 (●), pH 5 (□), and pH 2 (△). The changes in molar ellipticity at 230 nm and at 216 nm are plotted against pH in panels B and C, respectively. The solid line in panel B is a fit to eq 2 and gives apparent $pK_a$s of 6.3 and 3.7. The solid line in panel C is a fit to eq 1 and gives an apparent $pK_a$ of 3.5.

(1672 $cm^{-1}$), and a small decrease in the 1695 $cm^{-1}$ β-component.

At pH 2.0, the amide I spectrum of $I_U$ is different from that at pH 7.5 or 5.0, indicating that $I_U$ has a different secondary structure from native and $I_N$. The major changes are a large increase in the looplike structure at 1660 $cm^{-1}$ and loss of the major β-peak at 1638 $cm^{-1}$ in the native spectrum, which is replaced by a new, lower intensity β-component at 1631 $cm^{-1}$.

$^1$H NMR spectra were collected to further assess the conformational changes that took place in SMA at intermediate and low pH. As shown in Figure 8A, the NMR spectrum of SMA at pH 7.0 is characteristic of a tightly folded protein, having well-dispersed amide, aromatic, and aliphatic proton resonances. As the pH was reduced to below pH 5 (Figure 8B), only minor changes in the spectra were observed. These changes included both sharpening of many resonance lines as well as changes in some amide proton chemical shifts. However, the spectrum was still characteristic of a well-

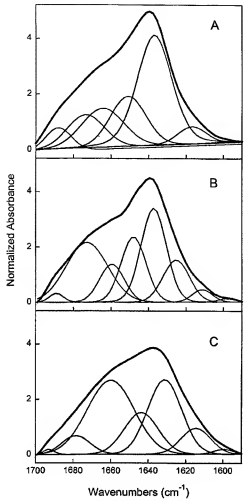



FIGURE 7: Amide I region of the FTIR spectrum of SMA. Panel A shows the spectrum for native SMA at pH 7.5. Panel B shows the spectrum of the nativelike intermediate ($I_N$) at pH 5.0, and panel C shows the spectrum of the unfolded-like intermediate ($I_U$) at pH 2.0. The raw ATR-FTIR spectra after water vapor subtraction are shown as thick lines. The thin lines in each panel are the component spectra obtained after curve-fitting the raw spectra.

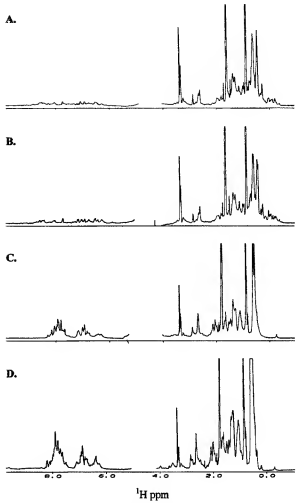



FIGURE 8: $^1$H NMR spectra of SMA. Panel A shows the spectrum of the native protein at pH 7. Panel B shows the spectrum of the nativelike intermediate, $I_N$, at pH 5, and panel C the spectrum of the relatively unfolded intermediate, $I_U$, at pH 2. For comparison, the spectrum of the unfolded protein in 8 M urea, pH 2, is shown in panel D.

folded protein. When the pH was adjusted to below 3, however, the chemical shift dispersion was lost, with the amide proton resonances collapsing to an envelope less than 1.5 ppm wide (Figure 8C). The upfield methyl resonances were also lost below pH 3, with both the aromatic and aliphatic proton regions of the spectrum showing considerable loss of dispersion. Notably, no resonances corresponding to either of the tryptophan indole moieties were apparent in the low-pH spectra. Additionally, the spectrum at pH 2.0 was considerably different from that recorded for SMA under strongly denaturing conditions (pH 2, 8 M urea) as shown in Figure 8D. Under these strongly denaturing conditions, a further loss of dispersion occurs throughout the spectrum, and significant changes in the chemical shifts of nearly all of the amide protons occur. As the solution pH is lowered, the NMR spectra show an increase in signal-to-noise for the same concentration of protein. This is likely due to dissociation of the $V_L$ dimer ($K_d = 40$ $\mu$M at pH 7) at lower pH.

*Small-Angle X-ray Scattering Characterization of SMA Conformations.* Small-angle X-ray scattering measurements indicated that SMA became less compact as the pH was reduced from 7 to 2 (at a protein concentration of 80 $\mu$M), with an increase in $R_g$ from 19.6 ± 0.4 Å at pH 7 to 26.8 ± 0.6 Å at pH 2. At pH 5, the protein was only slightly expanded, having a $R_g$ of 20.5 ± 0.6 Å. Kratky plots of the scattering data indicated that extensive globularity was maintained even at low pH, although some denaturation was apparent at pH 5 and more so at pH 2. The significant compactness of $I_U$ ($R_g = 26.8$ Å) is apparent by comparison of its $R_g$ to that of the fully unfolded protein ($R_g > 30$ Å).

*Equilibrium Thermodynamic Stability of SMA.* The stability of SMA was measured at different pHs using urea denaturation monitored by intrinsic tryptophan fluorescence and far-UV CD. Tryptophan fluorescence had the advantage that it permitted the use of low protein concentrations, which eliminated potential aggregation problems during unfolding. At pH 7.5, SMA is only marginally stable, with a free energy of unfolding $\Delta G° = 4.8$ kcal/mol and $m = 1.05$ kcal/mol. Similar equilibrium plots were obtained when starting with either native or unfolded protein, indicating that there was no hysteresis in the urea unfolding transitions. As the pH was decreased, the stability of SMA decreased significantly (Table 1) as indicated by the decrease in the midpoint of

Table 1: Effect of pH on the Stability of SMA[a]

| pH | $C_m$ (M) |
|---|---|
| 2 | ~0 |
| 4 | 1.5 ± 1.5 |
| 6 | 3.2 ± 0.1 |
| 8 | 4.0 ± 0.1 |
| 10 | 3.5 ± 0.1 |

[a] Midpoints of urea unfolding transitions were monitored by Trp fluorescence.

the urea unfolding transition with decrease in pH. In light of the evidence for formation of a non-native species from SMA in the vicinity of pH 4−6 ($I_N$), the urea unfolding data were not converted into free energy data except for neutral pH.

*Kinetics of Interconversion of Intermediates.* The rates of interconversion between the native conformation (N) and the partially folded intermediates ($I_N$ and $I_U$) were monitored by pH jumps. Interconversions between N and $I_N$ were monitored using changes in the ANS fluorescence and jumping the pH of solutions of SMA from 7 to 5 (N to $I_N$) and from 5 to 7 ($I_N$ to N). Interconversions involving $I_U$ were followed by observing changes in the tryptophan fluorescence, with pH jumps of 7 to 2 (N to $I_U$), 5 to 2 ($I_N$ to $I_U$), 2 to 5 ($I_U$ to $I_N$), and 2 to 7 ($I_U$ to N). The results from these experiments show that conversion of N to $I_N$ and $I_N$ to N are fast processes, complete within the dead time of manual mixing, indicating that the rate constant is >0.35 $s^{-1}$. On the other hand, conversion of either N or $I_N$ to $I_U$ and the reverse are slower processes. The rates for both N to $I_U$ and $I_N$ to $I_U$ were the same within experimental error, namely, 0.01 ± 0.002 $s^{-1}$, consistent with $I_N$ lying on the pathway between N and $I_U$. The rates of the transformation of $I_U$ to either $I_N$ or N were the same, with a rate constant of 0.002 ± 0.0008 $s^{-1}$, again consistent with $I_N$ lying on the pathway between N and $I_U$.

## DISCUSSION

There is increasing evidence to suggest that protein aggregation, including amyloid fibril formation, arises from a partially folded conformation of the aggregating protein (*5, 6, 26*−*30*). The present data strongly support such a hypothesis for the aggregation of immunoglobulin light chain variable domains. Protein aggregation has generally been regarded as being driven by nonspecific hydrophobic interactions operating on unfolded or collapsed molten-globule states. On the other hand, extracellular aggregation of some proteins to form amyloid fibrils has been conventionally attributed to mutations altering the local surface properties of the native state, thereby introducing new packing interactions for oligomerization of the native state (*26, 31*). In the case of SMA, our results clearly indicate that a nativelike conformation in the fibrils is highly unlikely. Thus, the model for light chain amyloid fibrils proposed by Stevens and coworkers (*31*), and based on the native structure, is inconsistent with the experimental observations.

The partially folded conformations that are the critical precursors to protein aggregation may arise either during the folding of newly synthesized proteins, as with inclusion bodies, or from the native state, as appears plausible for the extracellular amyloid deposits. The build-up of the soluble precursor that triggers aggregation could involve a combination of factors including an amino acid sequence leading to a relatively less stable native state as compared to that of a nonaggregating variant and/or the presence of mildly destabilizing extrinsic conditions.

Detailed analysis of the properties of the amyloidogenic SMA, including its stability, conditions necessary to populate partially folded intermediate conformations, propensity to aggregate, and kinetics of aggregation and fibril formation, provides insight into the molecular basis for aggregation. The present results raise a number of interesting questions, such as: Which features are responsible for the propensity to form fibrils? Why are both amorphous and fibrillar aggregates formed, and what is the relationship, if any, between them? How do the two partially folded intermediate conformations, $I_N$ and $I_U$, fit into the picture? We will begin with this last question.

*Destabilizing Conditions Lead to Partially Folded Intermediate Conformations.* Mildly destabilizing conditions, such as low pH or low urea concentrations (data not shown), lead to enhanced aggregation and fibrillation of SMA. The results of the spectroscopic investigations of SMA as a function of pH reveal that SMA forms two partially folded conformations, $I_N$ and $I_C$, the former being relatively nativelike in its structural properties, whereas the latter is considerably more unfolded. The major significance of the observation of these species is the correlation between formation of amorphous aggregates and $I_N$, and formation of fibrils and $I_U$. Both $I_N$ and $I_U$ are envisaged as ensembles of conformations that retain some nativelike structure (more in $I_N$ and less in $I_U$) with the remainder of the chain, especially for $I_U$, being highly mobile, and disordered but biased toward its native conformation.

The near- and far-UV circular dichroism, Trp fluorescence, NMR, and SAXS data all point to $I_N$ as being a relatively nativelike species with most structural properties similar to those of the native state. The significant increase in ANS binding, however, points to the critical feature of this intermediate, namely, increased exposure of hydrophobic surfaces compared to the native state. The increased negative ellipticity in the far-UV CD at 230 nm is related to this as it probably represents minor structural rearrangements in side-chain packing manifested as changes in the CD contribution of aromatic residues, rather than secondary structure changes. The enhanced ANS binding in the pH 4−6 range is very consistent with the population of a partially folded intermediate (*22, 23*). Likewise, the FTIR spectrum for $I_N$ reveals that although the overall secondary structure is quite similar to that of the native state, nevertheless there are significant structural differences. These include increased turn structure and a shift in some of the $\beta$-structure to components with a lower frequency band, perhaps signifying changes in the $\beta$-strand interactions. From examination of the spectral probes as a function of pH, it is apparent that there is a structural transition with a midpoint around pH 5.5. This transition, which corresponds to the interconversion of the native state to $I_N$, could reflect the titration of histidine or carboxylate residues in SMA. The pH-dependent transition from N to $I_N$ was not observed for the nonamyloidogenic LEN (Figure 5 and unpublished observations). The only differences in ionizable side chains between LEN and SMA are two histidines present in SMA (*8*); this suggests that one

or both of these His residues, either directly or indirectly, may be responsible for the transitions between N and $I_N$.

The data indicate that SMA forms a second, more unfolded, intermediate, $I_U$, at pH <3. From comparison of the spectral probe signals as a function of pH, a common transition, attributed to that of $I_N$ to $I_U$, with a midpoint in the vicinity of pH 3.3 is observed with all the probes. This intermediate retains substantial compactness, and secondary structure, consistent with the presence of a partially folded intermediate conformation. Based on the apparent p*K* of the transition between $I_N$ and $I_U$, this conformational change is apparently governed by the titration of carboxylate groups. The fact that at pH 2 the fluorescence signal reflects substantial residual structure (significantly decreased intensity and blue-shifted $\lambda_{max}$ relative to the unfolded and native states) suggests that there is also tertiary structure present in $I_U$. However, the near-UV CD spectrum indicates the loss of most of the native aromatic side-chain interactions, suggesting that the aromatic clusters present in the native conformation may no longer be present. The FTIR spectra indicate additional loop/disordered structure in $I_U$ compared to the native and $I_N$ conformations and also loss of nativelike $\beta$-structure. The NMR spectrum of this intermediate also clearly indicates that it retains considerable secondary and tertiary structure.

Among the factors that stimulate fibril formation from SMA are increased protein concentration, agitation, and increased temperature. All these are likely to result in increased concentration of non-native conformations (through equilibrium between native and non-native states, denaturation at air−water interfaces, and shifting of the equilibrium from native to non-native conformations, respectively), with their known propensity to aggregate. These observations strengthen the correlation between aggregation and precursor partially folded intermediates.

A minor complication in some of these experiments arises from the fact that $V_L$ domains are known to dimerize in a fashion similar to the association of the $C_L/V_L$ domain interaction in intact immunoglobulins. Stevens and Schiffer (*32*) demonstrated that native SMA exists in a monomer−dimer equilibrium with $K_d = 40$ $\mu$M under physiological conditions. Thus, data collected with high concentrations of SMA, such as SAXS and NMR, will be potentially complicated by the presence of these native dimers, at least at neutral pH. The main anticipated effect would be that under such conditions the equilibrium between the native conformation and the intermediates would be shifted in favor of the native conformation, rather than a non-native one. This leads to a potential decrease in the rate of fibril formation under conditions where the protein is present mostly as the native dimer (unpublished observations).

*Relation between Stability and Aggregation.* The stability of SMA is greatly decreased at acidic pH, correlating with both increased amorphous and fibrillar aggregation. The fact that at lower pH SMA readily forms fibrils suggests that removal of nativelike interactions is important prior to fibrillation. Our data show that the aggregation of the amyloidogenic SMA correlates well with the decreased stability of the native state and the population of non-native conformations. We believe that it is the *differential destabilization* of the native state of SMA relative to the partially folded intermediate conformations which is the key feature

Scheme 1

$$\begin{array}{ccccccc} N & \leftrightarrow & I_N & \leftrightarrow & I_U & \leftrightarrow & U \\ & & \downarrow & & \downarrow & & \\ & & \text{amorphous} & & \text{fibrils} & & \\ & & \text{aggregates} & & & & \end{array}$$

Scheme 2

$$\begin{array}{ccc} & & I_N \rightarrow \text{amorphous aggregates} \\ & \nearrow & \\ N & \leftrightarrow & U \\ & \searrow & \\ & & I_U \rightarrow \text{fibrils} \end{array}$$

of the amyloidogenesis, and which will be attenuated in nonamyloidogenic light chains. These observations are in accord with previous investigations of the correlation between amyloidogenesis and stability in light chains, which has indicated that either destabilizing mutations or sequences (*12, 33, 34*), or destabilizing conditions (*35*) correlate with increased fibillogenicity. In fact, SMA has been shown to be about 2.5 kcal/molecule less stable than its benign LEN homologue (*8*). Further, it has been shown that destabilizing LEN with GdnHCl leads to fibrillation (*12*).

*Amorphous and Fibrillar Aggregates.* The observation that both fibril and amorphous aggregate formation occur simultaneously under many conditions raises a number of questions, for example: Do they arise from similar or different partially folded conformations, and do amorphous aggregates convert directly or indirectly to fibrils?

The coincident kinetics for light scattering and TFT at pH ≤3, in conjunction with the limited amount of amorphous aggregates observed by EM, indicate that at these low pHs fibrils are preferentially formed, and that the limited amount of amorphous aggregates formed under these conditions rapidly converts to fibrillar species. Since the spectroscopic results suggest that at pH ≤3 the only significant species present is $I_U$, the limited amount of amorphous aggregates indicates that it is $I_U$ that is responsible for fibril formation. Similarly, the fact that the pH at which maximal amounts of amorphous material is found is in the vicinity of 5.5 indicates that $I_N$ is primarily responsible for the amorphous aggregates. The decreased amplitude of the final TFT signal at higher values of pH is attributed to the increased amount of amorphous aggregate at higher pHs.

The kinetics of interconversion between N, $I_N$, and $I_U$ are consistent with $I_N$ being on the pathway between N and $I_U$ (Scheme 1). However, the possibility that there are separate pathways to $I_N$ and $I_U$ (Scheme 2) cannot be eliminated at this time. The fast interchange between N and $I_N$ is not surprising, given the fact that $I_N$ is relatively nativelike.

The decreased amounts of fibrils observed at higher pH values presumably result from the smaller amounts of $I_U$ in equilibrium with N and $I_N$ (even though the equilibrium levels of $I_U$ may be quite low at higher pHs, any $I_U$ lost in fibril formation will be replaced by mass action with more soluble $I_U$). The strong correlation between the pH dependence of aggregation (and fibril formation) and the population of the partially folded intermediates supports the hypothesis that the observed intermediates are key players in the aggregation process. Fibril generation, even under nativelike conditions, can be attributed to the Boltzmann distribution of ensembles of various states under nativelike conditions. Hence, it possible that the key intermediate, highly populated

at pH ≤3, is also present under nativelike conditions, but at substantially lower concentration, and is responsible for amyloid formation after an extended lag period.

Aggregation results from the strong self-association tendency of the partially folded intermediates, probably due to the presence of large solvent-exposed hydrophobic patches, which are absent in the native and fully unfolded states. The increased $\beta$-structure observed in the aggregated states reflects $\beta$-strand−$\beta$-strand interactions involved in the intermolecular association. Fibril formation is expected to involve a number of intermediate states of soluble oligomers of partially folded intermediates, potentially populated at very low levels. Aggregation occurs under conditions in which a suitably high concentration of the key partially folded intermediate is present, due to a combination of destabilizing factors, such as pH and temperature, or urea, and amino acid sequence, as well as the concentration of the intermediate. Thus, it is mostly the intrinsically low stability of SMA that leads to the build-up of the intermediate leading to aggregation, under conditions where more stable light chains form negligible intermediate and remain in the native conformation.

The much more rapid formation of amorphous aggregates of SMA, compared to fibrils, is very interesting, and raises a number of questions regarding the nature of the relationship between the initially formed amorphous aggregates and the more slowly formed fibrils. Based on the data reported here, it is apparent that the partially folded intermediate populated in the pH 4−6 region is the direct precursor of the amorphous aggregates. The correlation between $I_N$ and amorphous aggregates, and $I_U$ and fibrillar aggregates, suggests that the ratio of the two types of deposits is determined, at least in part, by kinetic competition between the pathways leading to the two different intermediates. A more detailed investigation of the relationship between amorphous and fibrillar deposits will be given elsewhere.

The results of the present investigation firmly establish the existence of partially folded intermediates as key precursors on the aggregation pathway of the amyloidogenic light chain variable domain SMA. In addition, the observation of two such intermediates is the first report that a given protein might have more than one critical intermediate conformation on the aggregation pathway, and that such different conformations may lead to different types of deposits. One implication of this is that factors, such as chaperones, which may change the effective concentration of one of the intermediates may change the nature of the deposits.

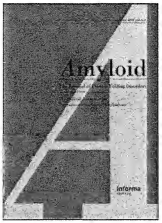
## ACKNOWLEDGMENT

We thank Vladimir N. Uversky and Pierre Soulliac for helpful discussions. Sue Carter for the atomic force microscope, and Sebastian Doniach for assistance with the small-angle X-ray scattering experiments.

## REFERENCES

1. Buxbaum, J. (1992) *Hematol. Oncol. Clin. North Am. 6*, 323−346.
2. Buxbaum, J., and Gallo, G. (1999) *Hematol. Oncol. Clin. North Am. 13*, 1235−1248.
3. Kaplan, B., Vidal, R., Kumar, A., Ghiso, J., Frangione, B., and Gallo, G. (1997) *Clin. Exp. Immunol. 110*, 472−478.
4. Buxbaum, J. N., Chuba, J. V., Hellman, G. C., Solomon, A., and Gallo, G. R. (1990) *Ann. Intern. Med. 112*, 455−464.
5. Lai, Z., Colón, W., and Kelly, J. W. (1996) *Biochemistry 35*, 6470−6482.
6. Booth, D. R., Sunde, M., Bellotti, V., Robinson, C. V., Hutchinson, W. L., Fraser, P. E., Hawkins, P. N., Dobson, C. M., Radford, S. E., Blake, C. C. F., and Pepys, M. B. (1997) *Nature 385*, 787−793.
7. Ionescu-Zanetti, C., Khurana, R., Gillespie, J. R., Petrick, J. S., Trabachino, L. C., Minert, L. J., Carter, S. A., and Fink, A. L. (1999) *Proc. Natl. Acad. Sci. U.S.A. 96*, 13175−13179.
8. Stevens, P. W., Raffen, R., Hanson, D. K., Deng, Y. L., Berrios-Hammond, M., Westholm, F. A., Murphy, C., Eulitz, M., Wetzel, R., Solomon, A., Schiffer, M., and Stevens, F. J. (1995) *Protein Sci. 4*, 421−432.
9. Pras, M., Schubert, M., Zucker-Franklin, D., Rimon, A., and Franklin, E. C. (1968) *J. Clin. Invest. 47*, 924−933.
10. Solomon, A. (1985) *Methods Enzymol. 116*, 101−121.
11. Huang, D. B., Chang, C. H., Ainsworth, C., Johnson, G., Solomon, A., Stevens, F. J., and Schiffer, M. (1997) *Mol. Immunol. 34*, 1291−1301.
12. Raffen, R., Dieckman, L. J., Szpunar, M., Wunschl, C., Pokkuluri, P. R., Dave, P., Wilkins, S. P., Cai, X., Schiffer, M., and Stevens, F. J. (1999) *Protein Sci. 8*, 509−517.
13. Skerra, A., Pfitzinger, I., and Pluckthun, A. (1991) *Biotechnology (N.Y.) 9*, 273−278.
14. Oberg, K. A., and Fink, A. L. (1998) *Anal. Biochem. 256*, 92−106.
15. Seshadri, S., Khurana, R., and Fink, A. L. (1999) *Methods Enzymol. 309*, 559−576.
16. Fink, A. L., Seshadri, S., Khurana, R., and Oberg, K. A. (1999) in *Infrared Analysis of Peptides and Proteins* (Singh, B. R., Ed.) American Chemical Society, Washington, DC.
17. Naiki, H., Higuchi, K., Hosokawa, M., and Takeda, T. (1989) *Anal. Biochem. 177*, 244−249.
18. Uversky, V. N., Karnoup, A. S., Khurana, R., Segel, D. J., Doniach, S., and Fink, A. L. (1999) *Protein Sci. 8*, 161−173.
19. Guinier, A., and Fournet, G. (1955) in *Small-angle scattering of X-rays*, p 268, Wiley, New York.
20. Levine, H. (1993) *Protein Sci. 2*, 404−410.
21. Jarrett, J. T., and Lansbury, P. T. (1992) *Biochemistry 31*, 12345−12352.
22. Goto, Y., and Fink, A. L. (1989) *Biochemistry 28*, 945−952.
23. Semisotnov, G. V., Rodionova, N. A., Razgulyaev, O. I., Uversky, V. N., Gripas, A. F., and Gilmanshin, R. I. (1991) *Biopolymers 31*, 119−128.
24. Fink, A. L. (1999) in *The Encyclopedia of Molecular Biology* (Creighton, T. E., Ed.) pp 140−142, John Wiley & Sons, New York.
25. Byler, D. M., and Susi, H. (1986) *Biopolymers 25*, 469−487.
26. Wetzel, R. (1996) *Cell 86*, 699−702.
27. Fink, A. L. (1998) *Folding Des. 3*, 9−15.
28. Speed, M. A., Wang, D. I. C., and King, J. (1996) *Nat. Biotechnol. 14*, 1283−1287.
29. Wetzel, R. (1999) *Amyloid, prions and other protein aggregates*, Academic Press, New York.
30. Quintas, A., Saraiva, M. J., and Brito, R. M. (1999) *J. Biol. Chem. 274*, 32943−32949.
31. Stevens, F. J., Myatt, E. A., Chang, C. H., Westholm, F. A., Eulitz, M., Weiss, D. T., Murphy, C., Solomon, A., and Schiffer, M. (1995) *Biochemistry 34*, 10697−10702.
32. Stevens, F. J., and Schiffer, M. (1995) *Methods Mol. Biol. 51*, 51−81.
33. Wall, J., Schell, M., Murphy, C., Hrncic, R., Stevens, F. J., and Solomon, A. (1999) *Biochemistry 38*, 14101−14108.
34. Bellotti, V., Stoppini, M., Mangione, P. P., Fornasieri, A., Minoletti, I., Merlini, G., and Ferri, G. (1996) *Biochim. Biophys. Acta 1317*, 161−167.
35. Kim, Y., Wall, J. S., Meyer, J., Murphy, C., Randolph, T. W., Manning, M. C., Solomon, A., and Carpenter, J. F. (2000) *J. Biol. Chem. 275*, 1570−1574.
36. Khurana, R., Hate, A. T., Nath, U., and Udgaonkar, J. B. (1995) *Protein Sci. 4*, 1133−1144.

BI001782B




**Amyloid**
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title-content=t713668970

# Protein and host factors implicated in the pathogenesis of light chain amyloidosis (AL amyloidosis)

Alan Solomon [a]; Deborah T. Weiss [a]
[a] Human Immunology & Cancer Program, Department of Medicine, University of Tennessee Medical Center, Graduate School of Medicine, Knoxville, TN, USA

**To cite this Article** Solomon, Alan and Weiss, Deborah T.(1995) 'Protein and host factors implicated in the pathogenesis of light chain amyloidosis (AL amyloidosis)', Amyloid, 2: 4, 269 — 279
**To link to this Article: DOI:** 10.3109/13506129508999010
**URL:** http://dx.doi.org/10.3109/13506129508999010

*Amyloid: Int. J. Exp. Clin. Invest. 2, 269-279 (1995)*

# Review

# Protein and host factors implicated in the pathogenesis of light chain amyloidosis (AL amyloidosis)

*Alan Solomon and Deborah T. Weiss*

Human Immunology & Cancer Program, Department of Medicine, University of Tennessee Medical Center, Graduate School of Medicine, Knoxville, TN, USA

*KEY WORDS: AMYLOIDOSIS, BENCE JONES PROTEINS, PLASMA CELL DYSCRASIA, IMMUNOGLOBULINS, LIGHT CHAINS*

*ABBREVIATIONS: AL: LIGHT CHAIN-ASSOCIATED AMYLOIDOSIS, IG: IMMUNOGLOBULIN, $V_L$: LIGHT CHAIN VARIABLE REGION, $C_L$: LIGHT CHAIN CONSTANT REGION, J: JOINING, CDR: COMPLEMENTARITY DETERMINING REGION, FR: FRAMEWORK REGION, GDNHCL: GUANIDINE HYDROCHLORIDE, IAPP: ISLET AMYLOID POLYPEPTIDE, PRP: PRION PROTEIN, AEF: AMYLOID ENHANCING FACTOR, IDOX: 4'-IODO-4-DEOXYDOXORUBICIN, GAG: GLYCOSAMINOGLYCAN, APO E: APOLIPOPROTEIN E*



## Introduction

Light chain-associated (AL) amyloidosis is characterized by the deposition in tissue of monoclonal light chain-related components – a pathologic process that leads to organ failure and eventual death. The lack of information on the etiology of this disease and the fact that there is no effective treatment to prevent or remove the abnormal tissue deposits represent major scientific challenges, the solutions of which may be relevant to other amyloid-associated disorders.

The purpose of this review is to summarize current research efforts that are directed towards elucidation of the protein and host factors thought to be responsible for the induction and perpetuation of AL amyloidosis. The insights that have emanated from these studies and the availability of new technologies, especially in structural and molecular biology, form the basis of future research. The ultimate goal is to apply this information clinically through design of novel therapeutic strategies that will be used to overcome the devastating impact of this disease.

### *Protein factors in the pathogenesis of AL amyloidosis*

Over 20 years have passed since Osserman and colleagues documented the almost invariable presence of serum or urinary monoclonal immunoglobulins (Igs) in patients diagnosed with what was previously termed primary amyloidosis[1,2]. The prediction that these molecules were involved in the pathogenesis of this disease[2] was confirmed when Glenner and co-workers (and subsequently, many other investigators) demonstrated that the amyloid deposits occurring in such patients were fibrillar in nature and composed of monoclonal light chains or, more commonly, light chain variable-region ($V_L$) fragments[3]. The unequivocal relationship between secreted and deposited light chains has since been established through comparative sequence analyses of Bence Jones proteins and amyloid proteins obtained from individual patients[3,4].

Historically, research on AL amyloidosis has been directed mainly towards characterizing the primary structural features of monoclonal light chain amyloid-associated proteins - i.e., those extracted from amyloid deposits or excreted in the form of Bence Jones protein, as

Correspondence: Dr. Alan Solomon, University of Tennessee Medical Center, 1924 Alcoa Highway, Knoxville, TN 37920, USA
Tel: 423-544-9165; Fax: 423-544-6865; e-mail: asolomon@wizard.hosp.utk.edu

Submitted: August 28, 1995 Revision Accepted: October 12, 1995

well as molecules encoded by DNA cloned from bone marrow-derived monoclonal plasma-cell populations. The goal of these studies has been to determine if particular chemical features of the light polypeptide chain render it amyloidogenic. A voluminous body of sequence data has been generated that describes apparently novel amino acid substitutions in the $V_L$ of such light chains[5]. Additionally, in some cases, post-translational modification (e.g., glycosylation) has been thought to account for this phenomenon[6]. However, given the rarity of carbohydrate-containing light chains and the fact that these components do not invariably form amyloid deposits, the functional or pathophysiological effects of glycosylation or other post-translational modifications (deamidation, etc.) are not known.

### *Light chain primary structure*

In an effort to differentiate amyloid proteins from those considered non-amyloidogenic, the primary structures of representative components have been compared. Unfortunately, these studies are confounded by the fact that while there is a relatively large referenced database on the former, the number of documented benign, "non-amyloid," or non-nephrotoxic light chains (Bence Jones proteins that are not associated with myeloma [east] nephropathy and light chain deposition disease) is markedly limited. Further, those proteins designated "non-amyloidogenic" may, in fact, represent amyloid-associated light chains derived from patients with unrecognized disease.

Another factor that complicates comparative analyses results from the variability inherent in light chain structure: First, there are two types of light chains, κ and λ, each having distinctive $V_L$ (and constant [$C_L$]) domains[5,7]. The $V_L$ portion of the molecule is the product of two exons – V and joining (J) – that encode, respectively, the first ~95 and following ~13 residues. The $V_\kappa$ and $V_\lambda$ domains are characterized by three hypervariable or complementarity-determining regions (CDRs) and four framework regions (FRs) that are involved, respectively, in the antigen-binding site and in maintaining the structural integrity of the molecule (Figure 1). Diversity in sequence results from the presence of multiple, functional $V_\kappa$- and $V_\lambda$- (~35 each) and $J_\kappa$- and $J_\lambda$- (5 and 4, respectively) germline genes, as well as the recombinatorial process that links the V and J segments (variation in the length of CDR3 at the V-J joint can also result from the presence at position 96 of non-template-encoded, extra residues that are the products of N or P nucleotides)[8-13].

Allelic differences can also account for diversity in primary structure, but of greater importance is the pronounced variability that is introduced in light chain-encoded germline sequences by somatic mutation. As yet, there has been no report of an amyloid associated light chain that is completely identical in sequence to that encoded by a $V_\kappa$- or $V_\lambda$-germline gene; however, the contributory role of somatic mutation to light chain amyloid formation remains to be established.

### *Structural features related to amyloidogenicity*

What are the structural features that account for the amyloidogenicity of certain light chains? It is noteworthy that in patients with AL amyloidosis λ-type monoclonal

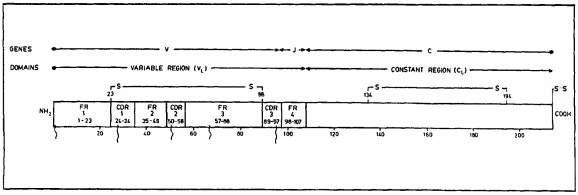



**FIGURE 1.** Schematic representation of $V_L$ and $C_L$ domains and their genetically-encoded segments. The $V_L$ is the product of two gene segments, V and J, that encode, respectively, the first ~95 residues and the remaining ~13 residues of the V region. The $C_L$ is the product of a single gene, designated C, that encodes for the remaining ~106 residues of the light chain. The location of the three hypervariable regions or CDRs and of the four FRs are as indicated; the J-encoded portion of the $V_L$ encompasses the terminal residue(s) of CDR3 and all the FR4 residues. The wavy lines symbolize the location of additional amino acid residues found within the FRs and CDRs. The residue numbering system is according to Reference 5.

Igs predominate, in contrast to the prevalence of κ-type components in other plasma-cell dyscrasias such as multiple myeloma, light chain deposition disease and Waldenström's macroglobulinemia[14]. Such overrepresentation of λ chains may be attributed to the fact that these components typically exist as covalent dimers, in contrast to κ molecules that occur in the monomeric as well as dimeric form[15]. This difference in quaternary structure may reduce the renal catabolism of λ-type components or, alternatively, increase their propensity to form amyloid. Most remarkable has been the finding that proteins of one particular Vλ subgroup – $V_{\lambda VI}$ – are invariably associated with this process[16]. Despite the relative rarity of this gene family in the normal λ-chain population (~5%), we have documented that 37% (13 of 35) of λ-type Igs obtained from patients with proven amyloidosis were of this isotype[17]. Among the primary structural features that distinguish these proteins are the presence of a unique two-residue insertion in FR3 that includes an Asp residue at position 66, as well as a Lys residue at position 17 in FR1. Although the tertiary structural features of λVI light chains have not as yet been elucidated, the interactive potential of these two oppositely charged amino acids may so alter the conformation of these molecules as to render them especially amyloidogenic (F. J. Stevens, personal communication).

To date, comparison of sequence data derived from amyloid and "non-amyloid" proteins (both λ- and κ-type) has failed to reveal a single, site-specific residue that differentiates between these two light chain populations. This finding (as well as the apparent irrelevance of glycosylation) differs from the case of the hemoglobinopathies where, in the case of hemoglobin S, for example, the one residue Glu/Val substitution in the β chain so alters the molecule that when deoxygenated sickling occurs[18].

In contrast to the theory that a single amino acid can render a light chain amyloidogenic, Stevens et al.[19] proposed an alternative mechanism to account for this phenomenon; namely, this process is based on a series of interactions between $V_L$-associated amino acid residues, first between those in monomers leading to the formation of dimers, then between dimers to form pre-amyloid filaments and finally, between filaments to form fibrils (Figure 2). Because multiple positions within the $V_L$ participate in β-domain interactions that result in $V_L$-dimer formation and ultimately amyloid fibrils (e.g., salt bridges, hydrogen bonds, or van der Waals contacts), almost any amino acid substitution could theoretically lead to an increase or decrease in the propensity of such interactions to occur, thereby modifying the tertiary structural features of the light chain and causing it, under appropriate physiological conditions, to become amyloidogenic. Conversely, particular amino acid substitutions could, in fact, inhibit light chain interactions and thus render the protein "non-amyloidogenic." The model satisfies several experimentally defined features of AL amyloid, including fibril diameter and the presence of crossed ß sheets that are aligned with the fibril axis and create an intrinsically self-stabilizing assembly.

This hypothesis was supported by the finding that amyloid-associated light chains contained potentially interactive, chemically distinct residues located at key positions within the $V_L$ domain[20]. A position-by-position comparison of the frequency of particular residues in 180 human monoclonal κ- and λ-type proteins, 52 of which were obtained from patients with known amyloidosis, revealed statistically significant differences in the CDRs and FRs of the two populations. For example, amyloid-associated molecules were most often characterized by the presence in the CDRs of sterically accessible, charged residues (e.g., Asp). Alternatively, in FR2, the typical basic residue found at position 45 in non-amyloid proteins was replaced by Asn and in FR1 at position 20, amyloid constituents had a hydrophobic Ile residue instead of Thr or Ser. For both κ and λ light chains, the majority of amyloid-associated positions were distributed along the surface of the light chain involved in the antigen-binding site and, therefore, would not participate in the internal packing of the $V_L$ domain. Further, the side chains of residues at such positions would extend into the solvent and be accessible for interaction with other molecules in solution.

These chemical differences that potentially modify light chain tertiary structure were also predicted to enhance $V_L$-dimer formation and result in protein aggregation. Earlier studies, using size-exclusion chromatography, demonstrated that, on the basis of polymer formation, Bence Jones proteins obtained from patients with amyloidosis or other types of pathologic light chain deposits could be distinguished from non-toxic components[21]. The demonstration that amyloid-associated proteins share unique primary structural features and readily aggregate in an *in vitro* system provides further evidence for the three-step molecular mechanism by which light chains form fibrils, i.e., dimer → filament → fibril[19,20].

Wetzel and colleagues[22] have hypothesized a somewhat different theory to explain light chain amyloidogenesis. They posited that, while multiple substitutions at key sites in the $V_L$ are responsible for light chain amyloidogenicity, this phenomenon of amyloid fibril formation results because such replacements so modify the tertiary structure that the protein becomes partially or completely unfolded. It is this intermediate form, then, that is the culprit in AL amyloid formation (as has been postulated to occur in the transthyretin-associated amyloidoses[23]). Experimental support for this theory was obtained using $V_L$ fragments

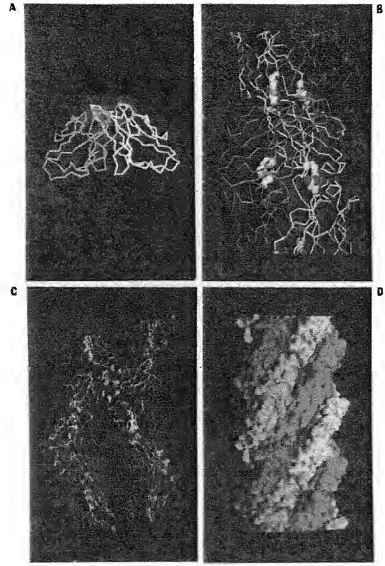



**FIGURE 2.** Computer-based models of $V_L$ interactions leading to the formation of amyloid fibrils. (A) Backbone representation of a $V_L$ dimer. This basic structure has been found for most Ig light chains which have been studied crystallographically. The shading surrounding the "backbone" indicates the volume occupied by atoms. The portion of the molecule that constitutes the equivalent of an antibody's antigen-binding site is located at the "top" of the molecule in this representation; among light chains, most amino acid substitutions occur in this area. (B) Hypothetical arrangement of $V_L$ dimers during the formation of non-covalent polymers. The antigen-binding site of one dimer interacts with the surface of the axially-opposed end of a second dimer following a 90° rotation about the two-fold symmetry axis. Many amyloid-forming light chains have amino acids of opposite charge, as highlighted in this figure, located in positions to form high-affinity "salt bridges." A consequence of the linking of $V_L$ domains is that any weak "non-specific" affinity of the dimer surface for particular tissue surface features, such as proteoglycans, is amplified. Thus, a light chain, which individually would not interact with an organ or tissue, might be specifically bound to an organ when polymerized. (C) Interaction of $V_L$-dimer filaments. When viewed from a different angle, the filament shown in (B) has a convoluted or "sawtooth" surface that makes it possible to bring together two or more filaments. In this figure, an anti-parallel mechanism of bringing filaments together is illustrated; several amino acid positions at which substitutions could either enhance or suppress the docking of filaments are highlighted. (D) Space-filling representation of the proposed $V_L$-dimer polymerization model. (Figure furnished by Dr. Fred J. Stevens, unpublished data.)

derived from DNA recombinant technology. These molecules, expressed in a bacterial system, were based on the $V_L$ dimer REI – a presumably non-amyloidogenic protein, the tertiary structure of which had been elucidated by x-ray crystallography[24]. Six $V_L$ constructs were prepared that contained replacements thought to be unique in κ- or λ-type amyloid-associated light chains. These modifications, due to their locations, were expected to destabilize the folded structure of the $V_L$ domain. In contrast to the wild-type REI component, all six mutants, at low pH, formed aggregates and bound Congo red. When examined in the presence of guanidine hydrochloride (GdnHCl), four of the six were unstable and aggregated to produce amyloid-like fibrils. The most destabilizing effect occurred when a highly-conserved Arg at position 61 was replaced by Asp. This substitution was predicted to eliminate or decrease the potential to form a salt bridge/hydrogen bond with another Asp residue located at position 82 in an adjacent loop. Although the GdnHCl in vitro stability assay may provide a means to predict the fibrillogenic capacity of a Bence Jones protein, the relevance of this system to *in vivo* amyloid formation remains to be established.

The aforementioned studies involving comparative analyses of light chain sequence have made it increasingly evident that such data *per se* are not sufficient to differentiate amyloid from non-amyloid components. In order to define more precisely the relationship between particular amino acid residues and light chain amyloidogenicity, it is necessary to determine how variations in sequence affect the overall tertiary structure of the molecule.

Because amyloid proteins cannot be crystallized, computer modeling offers a powerful new tool to predict the three-dimensional impact imparted on a protein by modification of and interactions between amino acids. Additionally, the ability to generate *via* recombinant technology a replenishable source of human light chains in which primary structural modification can be induced represents an important means to advance our understanding of the protein basis of amyloidogenicity[22,25,26]. This technique, as well as computer technology, provides exciting new approaches to research in this field, since manipulation of key amino acids and, most importantly, visualization of resulting interactions should provide information on why certain proteins form amyloid.

### *Light chain fragmentation in AL amyloidosis*

The structural requirements for the *in vivo* conversion of soluble light chain precursor proteins into insoluble amyloid fibrils have not been definitely established - e.g., it is unknown if the native protein itself can form amyloid fibrils *in vivo* or if this process is initiated or accelerated by light chain fragmentation. In rare instances, the protein extracted from AL amyloid deposits was found to be composed exclusively of an intact light chain[4,27]. More commonly, these extracts consist of light chain fragments ($V_L$ or $V_L$ plus a portion of the $C_L$)[3]. It has not been conclusively established if these components are produced *de novo*[28,29] or result from catabolism or *in situ* proteolysis of intact light polypeptide chains. The fact that we have found in virtually all AL amyloid extracts examined a certain, albeit small, amount of the complete light chain suggests that the common occurrence of fragments represents a degradative process. It is noteworthy that, based on the tertiary structural features of light chains, as evidenced by x-ray crystallography, the C-terminal residues of amyloid fragments (in positions 152-154) are typically located in sterically exposed regions of the molecule; such sites would thus be accessible for endopeptidase digestion. We have been able to generate similarly sized fragments using various types of endoproteases (e.g., cathepsin D, pepsin, or that contained in a lysosomal extract prepared from kidney). Additionally, we have found that amyloid fragments isolated from different organs from the same patient can vary in molecular mass[30]. Although such variations may reflect deposition of synthetically-derived light chain fragments, it is also possible that light chains are deposited in a relatively intact form as amyloid and that degradation occurs as mediated by local tissue factors. If light chain proteolysis is, indeed, essential for amyloid formation, the presence or extent of light chain fragmentation could have prognostic or therapeutic implications.

Experimentally, it has been shown that endoprotease digestion of certain Bence Jones proteins obtained from patients with or without amyloidosis yielded $V_L$ fragments that had the characteristic features of amyloid.[31-34] This finding may indicate that the $C_L$ portion of the molecule interferes with amyloid formation or, as has been shown *in vitro*, is remarkably susceptible to proteolysis as compared to the $V_L$ domain[35]. The demonstration that regions within the light chain are potentially amyloidogenic has come from studies of Eulitz and co-workers[36-38], who found that tryptic digestion of certain extensively reduced and alkylated amyloid-associated Bence Jones proteins yielded precipitates that exhibited green birefringence after Congo red-staining and were fibrillar when viewed by polarization light microscopy and electron microscopy, respectively. The insoluble tryptic digests, when rendered soluble in trifluoroacetic acid and subjected to HPLC-chromatographic separation on a reversed-phase column, were found to contain characteristic 18- to 43-residue $V_L$- and $C_L$-related peptides (this discovery was analogous to the observation that synthetic 10- to 28-residue peptides corresponding to portions of the Alzheimer's disease

amyloid-associated ß protein, the islet amyloid polypeptide [IAPP], or prion protein (PrP) could form amyloid fibrils *in vitro*[39-43]). More recent studies in our laboratory have led to the identification and synthesis of 12- to 24-mer $V_L$-related peptides that are particularly fibrillogenic (Ref. 30 and Niewold, Th. A. et al., unpublished studies). These findings suggest that, under appropriate conditions, all light chains are potentially amyloidogenic and that the susceptibility of these components to proteolytic cleavage may depend on other factors involved in light chain catabolism. That certain Bence Jones proteins can themselves function as peptidases[44] could also be of pathologic significance; however, the relevance of this phenomenon to the *in vivo* production of amyloid remains to be established.

If, indeed, 12- to 24-mer "amyloid" peptides are generated *in vivo* through proteolysis of precursor monoclonal light chains, such components may contribute to the pathogenesis of AL amyloidosis. Lansbury and colleagues[45-48] have proposed that amyloid formation requires or is accelerated by a seeding mechanism whereby preformed amyloid fibrils serve as a nucleation substance that initiates conversion of a soluble precursor protein or peptide into amyloid. Three factors were deemed essential for a nucleation-dependent polymerization process to occur. First, a lag time is necessary before detectable aggregates appear; second, there must be a critical concentration of the monomer before polymerization occurs; and third, polymerization is accelerated by a preformed "seed" or nucleus. That this phenomenon may ensue *in vivo* has been evidenced by the rapid formation of amyloid in the experimental hamster and mouse AA models[49,50] through the administration of "amyloid-enhancing factor" (AEF) - a sonicated preparation of an organ that contains either AA- or AL-related amyloid fibrils.

However, *in vitro* experiments have indicated that seeding is protein-specific; i.e., there must be complementarity between seed and some portion of the precursor protein for growth to occur, as shown in the case of ßA4 and prion fibrillogenesis[45,51-53]. The demonstration that amyloid formation can be accelerated by an apparent nucleation-dependent polymerization mechanism provides, as suggested by Jarrett and Lansbury[46], a rational therapeutic approach - namely, reduction of precursor protein concentration and interference with the generation of the nucleus should slow this relentless process. In this regard, $\alpha_1$-antichymotrypsin and a silicate compound ($Na_4SiO_4$) were found to inhibit ßA4 fibrillogenesis[54,55]. Another agent - 4'-iodo-4'-deoxydoxorubicin (IDOX), an iodinated anthracycline compound was shown to bind with high affinity to AL as well as other types of amyloid deposits and also interfered with *in vitro* fibrillogenesis. When incubated with AEF, this agent significantly reduced AA amyloid formation in an *in vivo* experimental animal system[56]. Further, five of eight patients with AL amyloidosis who were treated with this compound had objective evidence of amyloid resorption, despite no diminution in Bence Jones proteinuria[57].

### *Pathologic heterogeneity of light chain deposits*

Although the three major forms of the human light chain-associated diseases (myeloma [cast] nephropathy, light chain deposition disease and AL amyloidosis) exist as discrete entities[14,58,59], the coexistence of two distinct forms of light chain-associated disease - e.g., light-chain deposition disease and amyloidosis - has been noted[59-62]. It is not known whether each form results from the presence in individual patients of multiple Bence Jones protein components, from the potential of a single Bence Jones protein to induce more than one type of pathology, or from a natural progression of one disease state to another, as evidenced by the common involvement of blood vessel walls in non-fibrillar and fibrillar forms of light chain deposition.

The fact that a monoclonal light chain can assume multiple conformers according to the solvent used for crystallization[19,63] provides evidence that physiological factors can profoundly influence $V_L$ dimer interaction. The failure of a particular Bence Jones protein to form amyloid might be attributed to the greater propensity of the protein to aggregate as amorphous casts (myeloma [cast] nephropathy) or as punctate, linear deposits (light chain deposition disease). Alternatively (as previously discussed), particular amino acid substitutions may prevent $V_L$ aggregation or, if light chain fragmentation is a prerequisite for amyloid formation, such substitutions may so modify the native conformation of the light chain that proteasc-sensitive sites in the $C_L$ domain are rendered inaccessible. Furthermore, local factors may change the protein thus deposited: Initially the deposits may be non-fibrillar but are subsequently modified and assume the ß-pleated structure typical of amyloid fibrils.

## Host factors in the pathogenesis of AL amyloidosis Accessory Molecules

Interactions between monoclonal light chains and other biologically active molecules may also lead to pathologic deposition. For example, myeloma (cast) nephropathy has been thought to be the product of precipitated complexes comprised of Bence Jones protein and a component produced by the renal distal tubules – Tamm-Horsfall protein[64-66]. In the case of AL amyloid, the deposits, in addition to light chains, have been shown to contain several

chemically diverse substances, including amyloid P component[67], the highly-sulfated glycosaminoglycans (GAGs) heparan sulfate and dermatan sulfate[68-70] and apolipoprotein-E (Apo-E)[71]. Other types of molecules, such as ubiquitin, complement-related components, collagen, fibronectin and laminin, have been detected as well. The demonstration that certain of these compounds (e.g., Apo-E) can accelerate fibril formation in vitro has led to their designation as "pathologic chaperones[71-73]." It has been proposed that these molecules induce a ß-pleated-sheet conformation of the precursor protein that, in turn, leads eventually to fibril formation. However, we and others[23,34,36-38,74] have demonstrated that isolated light chains and related fragments can form amyloid fibrils *in vitro* in the absence of ancillary compounds. Thus, it is possible that the effects of these substances are secondary; i.e., binding is non-specific and results from the highly adsorptive nature of the ß-pleated-sheet structure of the amyloid fibril. The "pathologic chaperones," which are not found in non-amyloid forms of light chain deposition[62,72], may not necessarily facilitate fibrillogenesis but, rather, participate in other functions that are part of the amyloidogenic process – e.g., GAGs may play an important role in selective organ localization (see below) of AL amyloid deposits and P component has been implicated as a factor that can inhibit or prevent the resolution of this material[75].

### *Light chain metabolism*

Essential to the pathogenesis of AL amyloidosis is a continuous endogenous source of precursor protein, i.e., the monoclonal light chain. However, the amount of protein required for this process is unknown. Since the catabolic half-life of light chains is normally < 2 h, the concentration of these components in serum or urine reflects only a small percentage of the protein synthesized by the monoclonal plasma-cell population[76]. In contrast to patients with multiple myeloma, individuals with AL amyloidosis most often have a low percentage of bone marrow plasma cells (median 5%) and minimal Bence Jones proteinuria (median 0.4 g/24 h)[77]. Although there is evidence to suggest that a population of monoclonal B cells circulates and serves as precursor to the bone marrow-derived plasma cells[78], it is unlikely that there exists a significant extramedullary cellular source of monoclonal light chains (except in the case of localized AL amyloid deposits[79]). There is as yet no evidence to indicate that the light chain synthetic rate of monoclonal plasma cells obtained from patients with AL amyloidosis is greater than that of plasma cell populations associated with other forms of pathologic light chain deposition. The fact that patients with AL amyloidosis can have extensive disease despite a seemingly low serum or urine concentration of precursor proteins suggests (in the absence of an inordinately high rate of synthesis) an exceptional propensity of these molecules to form amyloid. It is also possible that such patients may have unusually high concentrations of light chain binding factors[80-82] or a molecular defect that limits or prevents the elimination of the AL deposits.

### *Organ diversity of AL deposits*

Central to the pathophysiology of this disease is the remarkable diversity that exists in the organs affected in patients with AL amyloidosis[77,83]. In some individuals, the deposits are confined principally to the kidney and in others to the heart, small intestine, or peripheral and/or sympathetic nerves, etc. Further, they may be primarily vascular, interstitial, or both. Our analyses of protein extracted from spleen and other organs have shown no relationship between the molecular mass of the deposited material and the affected tissue[30]. Whether selective tissue affinity is related to specific primary structural features of the light chain that result in an interaction with local tissue factors or to an antibody-like affinity of certain light chains for particular tissue constituents remains to be established. Previous attempts to demonstrate organ specificity *in vitro* using fluorescein-labeled native Bence Jones proteins obtained from patients with particular organ involvement, although showing qualitative differences, were considered inconclusive[1].

The demonstration by x-ray crystallography that Bence Jones proteins and $V_L$ dimers can structurally mimic Fab fragments, including the antigen-binding site[19], implies that the selective organ deposition of light chains may represent an antibody-ligand interaction. Alternatively, AL amyloid deposition may be due to the synthesis of the amyloidogenic precursor protein by a local monoclonal plasma-cell population. This phenomenon would be analogous to the site-specific production of other types of amyloid - e.g., ACal, AANF, AIAPP, Aß – associated with precursor proteins (calcitonin, atrial naturetic factor, islet amyloid polypeptide, ßA4 protein, respectively). Another mechanism that may account for the localization of AL amyloid deposits within particular organs involves the interaction between precursor light chains and tissue-specific molecules - e.g., GAGs. It remains to be determined if protein structure, organ affinity, or local production of light chains, etc. is responsible for the clinical diversity found in localized amyloid deposition.

### *Pathophysiology of AL amyloidosis*

Other physiologic factors analogous to those implicated in myeloma (cast) nephropathy may be involved in the pathogenesis of AL amyloidosis . For example, the precipitation of Bence Jones proteins in the form of renal

tubular casts is accelerated by dehydration or other conditions that can adversely affect renal function – e.g., hypercalcemia, anemia, or infection[14]. Due to the profound effect of solvent composition on light chain tertiary structure[63], it is possible that, under appropriate conditions, soluble precursor light chains can be induced to form amyloid fibrils.

Other host (patient)-related elements that could account for AL amyloid pathogenesis include the local processing by macrophages of soluble Bence Jones protein to form amyloid fibrils[84] or the failure of these cells to remove or prevent fibrillar light chain deposits. In contrast to the typical giant-cell reaction to Bence Jones protein-containing casts in myeloma (cast) nephropathy, characteristically, there is scant inflammatory response to AL amyloid deposits, presumably since this material is weakly or non-immunogenic[85,86]. With rare exception, AL amyloid deposits are apparently irreversible. The failure of the host to degrade AL amyloid fibrils may result from a deficiency in macrophage function or lack of a requisite proteolytic enzyme (as postulated in the pathogenesis of brain amyloid in Alzheimer's disease[87,88]). The coating of amyloid fibrils by "pathologic chaperones" - e.g., P component[75] – may also interfere with fibril degradation. Although there is no documented difference in the association of such components with λ- or κ-type amyloid fibrils, it has been reported that α1-antitrypsin was capable of disaggregating λ- but not κ-chain-containing fibrils[89].

### *In Vivo models of AL amyloidosis*

Further insight into the pathogenesis and, ultimately, the effective treatment of AL amyloidosis depends, in part, on the availability of *in vivo* experimental models that duplicate the human form of this disease. We have reported the results obtained with one such model whereby mice were repeatedly injected with different Bence Jones proteins obtained from patients with AL amyloidosis. In these experiments, the human proteins were deposited in the kidney and other organs of the recipient animals in the form of amyloid[90,91]. The human light chain amyloid also contained mouse amyloid P component. Conversely, no deposits were found in mice injected under similar conditions with a "control," i.e., non-amyloid, Bence Jones protein. Due to the relatively large amount of protein needed for injection and the length of time required for the amyloid to form, other models in which human amyloid-associated light chains are produced trangenically or by transfectomas may prove more suitable. The development of light chain-related amyloid in such *in vivo* models will provide an invaluable means for further research on the pathogenesis, treatment and prevention of AL amyloidosis.

## Acknowledgments

The helpful discussions and research contributions of our collaborating scientists, Dr. Fred J. Stevens (who also supplied Figure 2), Dr. Manfred Eulitz and Dr. Theo A. Niewold, are gratefully acknowledged. We thank Julie Ottinger for manuscript preparation. This study was supported in part by U.S. Public Health Research Grant CA10056 from the National Cancer Institute and the Stein Cancer Research Fund. A.S. is an American Cancer Society Clinical Research Professor.

## References

1 Osserman EF, Takatsuki K and Talal N (1964). Multiple myeloma. I. The pathogenesis of "amyloidosis." *Semin Hematol* **1**, 3-86

2 Isobe T and Osserman EF (1974). Patterns of amyloidosis and their association with plasma-cell dyscrasia, monoclonal immunoglobulins and Bence-Jones proteins. *N Engl J Med* **290**, 473-477

3 Glenner GG (1980). Amyloid deposits and amyloidosis: The β-fibrilloses. *N Engl J Med* **302**, 1283-1292; 1333-1343

4 Klafki H-W, Kratzin HD, Pick AI, Eckart K, Karas M and Hilschmann N (1992). Complete amino acid sequence determinations demonstrate identity of the urinary Bence Jones protein (BJP-DIA) and the amyloid fibril protein (AL-DIA) in a case of AL-amyloidosis. *Biochemistry* **31**, 3265-3272

5 Kabat EH, Wu TT, Perry HM, Gottesman KS and Foeller C (eds.) (1991). *Sequences of Proteins of Immunological Interest*, volume 1, 5th edition. (U.S. Department of Health and Human Services, PHS, NIH)

6 Nyquist J, Ramstad HM, Sletten K, Husby G and Westermark P (1994). Structural studies of two carbohydrate-containing immunoglobulin-kappa-light chain amyloid fibril-proteins of the variable subgroup I. In Kisilevsky R, Benson MD, Frangione B, Gauldie J, Muckle TJ and Young ID (eds.) *Amyloid and Amyloidosis* 1993, pp. 247-249. (New York: Partheon Publishing)

7 Stewart AK and Schwartz RS (1994). Immunoglobulin V regions and the B cell. *Blood* **83**, 1717-1730

8 Klein R, Jaenichen R and Zachau HG (1993). Expressed human immunoglobulin κ genes and their hypermutation. *Eur J Immunol* **23**, 3248-3271

9 Frippiat J-P and Lefranc M-P (1994). Genomic organisation of 34 kb of the human immunoglobulin lambda locus (IGLV): Restriction map and sequences of new VλIII genes. *Mol Immunol* **31**, 657-670

10 Bauer TR Jr and Blomberg B (1991). The human λ L chain Ig locus. Recharacterization of JCλ6 and identification of a

functional JCλ7. *J Immunol* **146,** 2813-2820

11 Hieter PA, Maizel JV Jr and Leder P (1982). Evolution of human immunoglobulin κ J region genes. *J Biol Chem* **257,** 1516-1522

12 Lafaille JJ, DeCloux A, Bonneville M, Takagaki Y and Tonegawa S (1989). Junctional sequences of T cell receptor γδ genes: Implications for γδ T cell lineages and for a novel intermediate of V-(D)-J joining. *Cell* **59,** 859-870

13 Victor KD and Capra JD (1994). An apparently common mechanism of generating antibody diversity: Length variation of the VL-JL junction. *Mol Immunol* **31,** 39-46

14 Solomon A (1986). Clinical implications of monoclonal light chains. *Semin Oncol* **13,** 341-349

15 Bernier GM and Putnam FW (1963). Monomer-dimer forms of Bence Jones proteins. *Nature* **200,** 223-225

16 Solomon A, Frangione B and Franklin EC (1982). Bence Jones proteins and light chains of immunoglobulins. Preferential association of the $V_{\lambda VI}$ subgroup of human light chains with amyloidosis AL(l). *J Clin Invest* **70,** 453-460

17 Ozaki S, Abe M, Wolfenbarger D, Weiss DT and Solomon A (1994). Preferential expression of human λ-light-chain variable-region subgroups in multiple myeloma, AL amyloidosis and Waldenström's macroglobulinemia. *Clin Immunol Immunopathol* **71,** 183-189

18 Bookchin RM and Nagel RL (1974). Interactions between human hemoglobins: Sickling and related phenomena. *Semin Hematol* **11,** 577-595

19 Stevens FJ, Solomon A and Schiffer M (1991). Bence Jones proteins: A powerful tool for the fundamental study of protein chemistry and pathophysiology. *Biochemistry* **30,** 6803-6805

20 Stevens FJ, Myatt EA, Chang C-H, Westholm FA, Eulitz M, Weiss DT, Murphy C, Solomon A and Schiffer M (1995). A molecular model for self-assembly of amyloid fibrils: Immunoglobulin light chains. *Biochemistry* **34,** 10697-10702

21 Myatt EA, Westholm FA, Weiss DT, Solomon A, Schiffer M and Stevens FJ (1994). Pathogenic potential of human monoclonal immunoglobulin light chains: Relationship of *in vitro* aggregation to *in vivo* organ deposition. *Proc Natl Acad Sci USA* **91,** 3034-3038

22 Hurle MR, Helms LR, Li L, Chan W and Wetzel R (1994). A role for destabilizing amino acid replacements in light-chain amyloidosis. *Proc Natl Acad Sci USA* **91,** 5446-5450

23 Kelly JW and Lansbury PT Jr (1994). A chemical approach to elucidate the mechanism of transthyretin and β-protein amyloid fibril formation. *Amyloid: Int J Exp Clin Invest* **1,** 186-205

24 Epp O, Lattman EE, Schiffer M, Huber R and Palm W (1975). The molecular structure of a dimer composed of the variable portions of the Bence-Jones protein REI refined at 2.0-Å resolution. *Biochemistry* **14,** 4943-4952

25 Wilkins Stevens P, Raffen R, Hanson DK, Deng Y-L, Berrios-Hammond M, Westholm FA, Murphy C, Eulitz M, Wetzel R, Solomon A, Schiffer M and Stevens FJ (1995). Recombinant immunoglobulin variable domains generated from synthetic genes provide a system for *in vitro* characterization of light-chain amyloid proteins. *Protein Sci* **4,** 421-432

26 Schormann N, Murrell JR, Liepnieks JJ and Benson MD (1995). Tertiary structure of an amyloid immunoglobulin light chain protein: A proposed model for amyloid fibril formation. *Proc Natl Acad Sci USA* **92,** 9490-9494

27 Terry WD, Page DL, Kimura S, Isobe T, Osserman EF and Glenner GG (1973). Structural identity of Bence Jones and amyloid fibril proteins in a patient with plasma cell dyscrasia and amyloidosis. *J Clin Invest* **52,** 1276-1281

28 Buxbaum J (1986). Aberrant immunoglobulin synthesis in light chain amyloidosis. Free light chain and light chain fragment production by human bone marrow cells in short-term tissue culture. *J Clin Invest* **78,** 798-806

29 Preud'homme JL, Ganeval D, Grünfeld JP, Striker L and Brouet JC (1988). Immunoglobulin synthesis in primary and myeloma amyloidosis. *Clin Exp Immunol* **73,** 389-394

30 Solomon A and Weiss DT (1994). AL amyloidosis. In Kisilevsky R, Benson MD, Frangione B, Gauldie J, Muckle TJ and Young ID (eds.) *Amyloid and Amyloidosis* 1993, pp. 200-205. (New York: Parthenon Publishing)

31 Glenner GG, Ein D, Eanes ED, Bladen HA, Terry W and Page DL (1971). Creation of "amyloid" fibrils from Bence Jones proteins in vitro. *Science* **174,** 712-714

32 Linke RP, Tischendorf FW, Zucker-Franklin D and Franklin EC (1973). The formation of amyloid-like fibrils in vitro from Bence Jones proteins of the VλI subclass. *J Immunol* **111,** 24-26

33 Glenner GG, Eanes ED, Bladen HA, Linke RP and Termine JD (1974). β-pleated sheet fibrils. A comparison of native amyloid with synthetic protein fibrils. *J Histochem Cytochem* **22,** 1141-1158

34 Epstein WV, Tan M and Wood IS (1974). Formation of "amyloid" fibrils in vitro by action of human kidney lysosomal enzymes on Bence Jones proteins. *J Lab Clin Med* **84,** 107-110

35 Solomon A and McLaughlin CL (1969). Bence-Jones proteins and light chains of immunoglobulins. I. Formation and characterization of amino-terminal (variant) and carboxyl-terminal (constant) halves. *J Biol Chem* **244,** 3393-3404

36 Eulitz M and Linke R (1985). Amyloid fibrils derived from V-region together with C-region fragments from a λII-immunoglobulin light chain (HAR). *Biol Chem Hoppe-Seyler* **366,** 907-915

37 Eulitz M, Breuer W and Linke RP (1987). Is the formation of AL-type amyloid promoted by structural peculiarities of

immunoglobulin L-chains? *Biol Chem Hoppe-Seyler* **368**, 863-870

38 Eulitz M, Breuer M, Ehlen A, Weiss DT and Solomon A (1990). Production of amyloidogenic peptides from human immunoglobulin light (L)-chains. In Natvig JB, Førre Ø, Husby G, Husebekk A, Skogen B, Sletten K and Westermark P (eds.) *Amyloid and Amyloidosis* 1990, pp. 505-510. (Dordrecht, The Netherlands: Kluwer Academic Publishers

39 Castaño EM, Ghiso J, Prelli F, Gorevic PD, Migheli A. and Frangione B (1986). In vitro formation of amyloid fibrils from two synthetic peptides of different lengths homologous to Alzheimer's disease ß-protein. *Biochem Biophys Res Commun* **141**, 782-789

40 Gorevic PD, Castaño EM, Sarma R and Frangione B (1987). Ten to fourteen residue peptides of Alzheimer's disease are sufficient for amyloid fibril formation and its characteristic Xray diffraction pattern. *Biochem Biophys Res Commun* **147**, 854-862

41 Kirschner DA, Inouye H, Duffy LK, Sinclair A, Lind M and Selkoe DJ (1987). Synthetic peptide homologous to ß protein from Alzheimer disease forms amyloid-like fibrils *in vitro*. *Proc Natl Acad Sci USA* **84**, 6953-6957

42 Westermark P, Engström U, Johnson KH, Westermark GT and Betsholtz C (1990). Islet amyloid polypeptide: Pinpointing amino acid residues linked to amyloid fibril formation. *Proc Natl Acad Sci USA* **87**, 5036-5040

43 Tagliavini F, Prelli F, Verga L, Giaccone G, Sarma R, Gorevic P, Ghetti B, Passerini F, Ghiboudi E, Forloni G, Salmona M, Bugiani O and Frangione B (1993). Synthetic peptides homologous to prion protein residues 106-147 form amyloid-like fibrils *in vitro*. *Proc Natl Acad Sci USA* **90**, 9678-9682

44 Paul S, Li L, Kalaga R, Wilkins-Stevens P, Stevens FJ and Solomon A (1995). Natural catalytic antibodies: Peptide-hydrolyzing activities of Bence Jones proteins and $V_L$ fragment. *J Biol Chem* **270**, 15257-15261

45 Jarrett JT and Lansbury PT Jr (1992). Amyloid fibril formation requires a chemically discriminating nucleation event: Studies of an amyloidogenic sequence from the bacterial protein OsmB. *Biochemistry* **31**, 12345-12352

46 Jarrett JT and Lansbury PT Jr (1993). Seeding "one-dimensional crystallization" of amyloid: A pathogenic mechanism in Alzheimer's disease and scrapie? *Cell* **73**, 1055-1058

47 Jarrett JT, Berger EP and Lansbury PT Jr (1993) The carboxy terminus of the ß amyloid protein is critical for the seeding of amyloid formation: Implications for the pathogenesis of Alzheimer's disease. *Biochemistry* **32**, 4693-4697

48 Come JH, Fraser PE and Lansbury PT Jr (1993). A kinetic model for amyloid formation in the prion diseases: Importance of seeding. *Proc Natl Acad Sci USA* **90**:5959-5963

49 Niewold TA, Hol PR, van Andel ACJ, Lutz ETG and Gruys E (1987) Enhancement of amyloid induction by amyloid fibril fragments in hamster. *Lab Invest* **56**, 544-549

50 Cathcart ES (1995). AEF: An enigma wrapped in a mystery. *Amyloid: Int J Exp Clin Invest* **2**, 126-127

51 Niewold TiA, Gruys E, Kisilevsky R and Shirahama TS (1991). Fibril amyloid enhancing factor (FAEF)-accelerated amyloidosis in the hamster is not dependent on scrine esterase activity and mononuclear phagocytosis. *Scand J Immunol* **34**, 101-107

52 Varga J, Flinn MSM, Shirahama T, Rodgers OG and Cohen AS (1986). The induction of accelerated murine amyloid with human splenic extract. *Virchows Arch* [B] **51**, 177-185

53 Ali-Khan ZI, Quirion R, Rabitaille Y, Alizadeh-Khiavi K and Du T (1988). Evidence for amyloid enhancing factor in Alzheimer brain extract. *Acta Neuropathol* **77**, 82-90.

54 Eriksson S, Janciauskiene S and Lannfelt L (1995). $\alpha_1$-antichymotrypsin regulates Alzheimer ß-amyloid peptide fibril formation. *Proc Natl Acad Sci USA* **92**, 2313-2317

55 Fasman GD, Perczel A and Moore CD (1995). Solubilization of ß-amyloid-(1-42)-peptide: Reversing the ß-sheet conformation induced by aluminum with silicates. *Proc Natl Acad Sci USA* **92**, 369-371

56 Merlini G, Ascari E, Amboldi N, Bellotti V, Arbustini E, Perfetti V, Ferrari M, Zorzoli I, Marinone MG, Garini P, Diegoli M, Trizio D and Ballinari D (1995) Interaction of the anthracycline 4'-iodo-4'-deoxydoxorubicin with amyloid fibrils: Inhibition of amyloidogenesis. *Proc Natl Acad Sci USA* **92**, 2959-2963

57 Gianni L, Bellotti V, Gianni AM and Merlini G (1995). New drug therapy of amyloidoses: Resorption of AL-type deposits with 4'-iodo-4'-deoxydoxorubicin. *Blood* **86**, 855-861

58 Sanders PW, Herrera GA, Kirk KA, Old CW and Galla JH (1991). Spectrum of glomerular and tubulointerstitial renal lesions associated with monotypical immunoglobulin light chain deposition. *Lab Invest* **64**, 527-537

59 Ganeval D, Noël L-H, Preud'homme J-L, Droz D and Grünfeld J-P (1984). Light-chain deposition disease: Its relation with AL-type amyloidosis. *Kidney Int* **26**, 1-9

60 Jacquot C, Saint-Andre J-P, Touchard G, Nochy D, D'Auzac de Lamartinie C, Oriol R, Druet P and Bariety J (1985). Association of systemic light-chain deposition disease and amyloidosis: A report of three patients with renal involvement. *Clin Nephrol* **24**, 93-98

61 Troussard X, de Ligny BH, Gallet B, Ganeval D, Mandard J-C, Ryckelynck J-P and Leporrier M (1989). Massive systemic amyloidosis associated with light-chain deposition disease. *Nephron* **52**,139-143

62 Buxbaum JN, Chuba JV, Hellman GC, Solomon A and Gallo GR (1990). Monoclonal immunoglobulin deposition disease: Light chain and heavy chain deposition diseases

and their relation to light chain amyloidosis. *Ann Intern Med* **112**, 455-464

63 Chang C-H, Short MT, Westholm FA, Stevens FJ, Wang B-C, Furey W Jr, Solomon A and Schiffer M (1985). Novel arrangement of immunoglobulin variable domains: X-ray crystallographic analysis of the λ-chain dimer Bence-Jones protein Loc. *Biochemistry* **24**, 4890-4897

64 Sanders PW, Herrera GA and Galla JH (1987). Human Bence Jones protein toxicity in rat proximal tubule epithelium in vivo. *Kidney Int* **32**, 851-861

65 Sanders PW, Booker BB, Bishop JB and Cheung HC (1990). Mechanisms of intranephronal proteinaceous cast formation by low molecular weight proteins. *J Clin Invest* **85**, 570-576

66 Sanders PW and Booker BB (1992). Pathobiology of cast nephropathy from human Bence Jones proteins. *J Clin Invest* **89**, 630-639

67 Baltz ML, Caspi D, Evans DJ, Rowe IF, Hind CRK and Pepys MB (1986). Circulating serum amyloid P component is the precursor of amyloid P component in tissue amyloid deposits. *Clin Exp Immunol* **66**, 691-700

68 Nelson ST, Lyon M, Gallagher JT, Johnson EA and Pepys MB (1991). Isolation and characterization of the integral glycosaminoglycan constituents of human amyloid A and monoclonal light-chain amyloid fibrils. *Biochem J* **275**, 67-73

69 Stenstad T, Magnus JH, Husby G and Kolset SO (1993). Purification of amyloid-associated heparan sulphate proteoglycans and galactosaminoglycan free chains from human tissues. *Scand J Immunol* **37**, 227-235

70 Snow AD, Willner J and Kisilevsky R (1987). Sulphated glycosaminoglycans: A common constituent of all amyloid? *Lab Invest* **56**, 120-124

71 Wisniewski T and Frangione B (1992). Apolipoprotein E: A pathological chaperone protein in patients with cerebral and systemic amyloid. *Neurosci Lett* **135**, 235-238

72 Gallo G, Wisniewski T, Choi-Miura N-H, Ghiso J and Frangione B (1994). Potential role of apolipoprotein-E in fibrillogenesis. *Am J Pathol* **145**, 526-530

73 Frangione B, Wisniewski T and Ghiso J (1994). Chaperoning Alzheimer's amyloids. *Neurobiol Aging* **15**, S97-S99

74 Klafki H, Pick AI, Pardowitz I, Cole T, Awni LA, Barnikol H, Mayer F, Kratzin HD and Hilschmann N (1993). Reduction of disulfide bonds in an amyloidogenic Bence Jones protein leads to formation of "amyloid-like" fibrils in vitro. *Biol Chem Hoppe-Seyler* **374**, 1117-1122

75 Tennent GA, Lovat LB and Pepys MB (1995). Serum amyloid P component prevents proteolysis of the amyloid fibrils of Alzheimer disease and systemic amyloidosis. *Proc Natl Acad Sci USA* **92**, 4299-4303

76 Solomon A, Waldmann TA, Fahey JL and McFarlane AS (1964). Metabolism of Bence Jones proteins. *J Clin Invest* **43**, 103-117

77 Kyle RA and Gertz MA (1995). Primary systemic amyloidosis: Clinical and laboratory features in 474 cases. *Semin Hematol* **32**, 45-59

78 Perfetti V, Bellotti V, Garini P, Zorzoli I, Rovati B, Marinone MG, Ippoliti G and Merlini G. (1994) AL amyloidosis. Characterization of amyloidogenic cells by anti-idiotypic monoclonal antibodies. *Lab Invest* **71**, 853-861

79 Miura K and Shirasawa H (1993). Lambda III subgroup immunoglobulin light chains are precursor proteins of nodular pulmonary amyloidosis. *Am J Clin Pathol* **100**, 561-566

80 Batuman V, Dreisbach AW and Cyran J (1990). Light-chain binding sites on renal brush-border membranes. *Am J Physiol* **258**, F1259-F1265

81 Hendershot LM (1990). Immunoglobulin heavy chain and binding protein complexes are dissociated in vivo by light chain addition. *J Cell Biol* **111**, 829-837

82 Haas IG (1991). BiP–A heat shock protein involved in immunoglobulin chain assembly. *Curr Top Microbiol Immunol* **167**, 71-82

83 Stone MJ (1990). Amyloidosis: A final common pathway for protein deposition in tissues. *Blood* **75**, 531-545

84 Durie BGM, Persky B, Soehnlen BJ, Grogan TM and Salmon SE (1982) Amyloid production in human myeloma stem-cell culture, with morphologic evidence of amyloid secretion by associated macrophages. *N Engl J Med* **307**, 1689-1692

85 Franklin EC and Pras M (1969). Immunologic studies of water-soluble amyloid fibrils. Comparative studies of eight amyloid preparations. *J Exp Med* **130**, 797-808

86 Benson M, Skinner M and Cohen AS (1975). Antigenicity and cross-reactivity of denatured fibril proteins of primary, secondary and myeloma associated amyloids. *J Lab Clin Med* **85**, 650-659

87 Glenner GG (1988). Alzheimer's disease: Its proteins and genes. *Cell* **52**, 307-308

88 Abraham CR, Selkoe DJ and Potter H (1988). Immunochemical identification of the serine protease inhibitor $\alpha_1$-antichymotrypsin in the brain amyloid deposits of Alzheimer's disease. *Cell* **52**, 487-501

89 Eriksson S, Janciauskiene S and Merlini G (1995). The putative role of alpha-1-antitrypsin in the disaggregation of amyloid λ fibrils. *J Int Med* **237**, 143-149

90 Solomon A, Weiss DT and Kattine AA (1991). Nephrotoxic potential of Bence Jones proteins. *N Engl J Med* **324**, 1845-1851

91 Solomon A, Weiss DT and Pepys MB (1992). Induction in mice of human light-chain-associated amyloidosis. *Am J Pathol* **140**, 629-637

# Introduction to the Membrane Protein Reviews: The Interplay of Structure, Dynamics, and Environment in Membrane Protein Function

Jonathan N. Sachs and Donald M. Engelman

Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520; email: jsachs@csb.yale.edu, donald.engelman@yale.edu

Annu. Rev. Biochem.
2006. 75:707–12

First published online as a Review in Advance on March 20, 2006

The *Annual Review of Biochemistry* is online at biochem.annualreviews.org

doi: 10.1146/annurev.biochem.75.110105.142336



0066-4154/06/0707-0707$20.00

## Abstract

In our review, we introduce an organizational scheme for membrane protein function. It is the relationship between structure, dynamics, and environment that endows the membrane and its constituents with remarkable sensitivity and robustness. Our understanding begins with landmark advances like those presented in the following chapters. Membrane proteins are notoriously difficult to study, and so the work presented here on the ADP/ATP carrier [Nury et al. (2)], rhodopsin [Palczewski (24)], and the cytochrome $b_6f$ complex [Cramer et al. (35)] represents incredible progress in this now blossoming field.

## INTRODUCTION

Membrane protein structure determination is proceeding at an exciting pace, driven by the hope that structures can connect decades of biochemical and biophysical observations to protein function and mechanism. The reviews that follow demonstrate that structures are capable of resolving long-standing mysteries while suggesting new questions and opening new areas of scientific discovery. As science as a whole unravels increasingly complex and finely tuned functionality, can structure determination continue to keep pace, or will new methodologies and perspectives be required? The ultimate test will be to build robust molecular-level models capable of predicting the functional outcome of any arbitrary structural perturbation. The creation of such models will require adopting a broad perspective, one that takes into account dynamic deviations from static structures as well as the influence of the membrane environment.

It is helpful to establish a paradigm that categorizes and organizes the relevant contributors to function. Structures should, and will, act as the nucleation points for this organization. This is because understanding mechanisms is ultimately a matter of chemistry, and one cannot invoke chemical ideas of the mechanism without knowing where the atoms are. As we detail in the discussion that follows, evolution has exploited three categories of molecular-level organization in order to achieve efficient and diverse membrane protein functions: structure, molecular dynamics, and environmental constraints. As suggested schematically in **Figure 1**, they are inextricably linked, each influencing the other, collectively dictating membrane protein function. By recasting the structure-function relationship in this way, we suggest that a comprehensive view of membrane protein function may be more readily achieved.

## STRUCTURE

Any modern model for predicting function must start with structure. As demonstrated in the following reviews, static structures can elegantly illuminate the structure-function relationship. Collectively, the articles detail a series of new features only now visible in the increasingly high-resolution structures. The emergence of these elements, which include protein oligomerization and complexation, lipidation, glycosylation, and the presence of structural waters, suggests that future models may need to account for a high level of complexity and structural variability.

Oligomerization is increasingly seen as a common motif in membrane proteins (1), and dimerization in particular is thought to be the rule for each of the three proteins discussed below. As the authors point out, distinguishing between dimers and higher order oligomers is far from trivial. Although this difficulty may be primarily due to experimental limitations such as dissociation by detergent, it may also reflect true biological variability: Perhaps there is a functionally

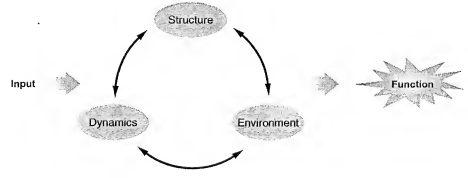



Figure 1
The three categories of molecular-level organization needed to achieve efficient and diverse membrane protein functions.

relevant equilibrium distribution of *n*-mers within cell membranes. This possibility suggests an investigation into the evolutionary origin of oligomerization. We suggest four distinct, though not exclusive, explanations. First, in most studied cases, oligomerization serves a functional role, and this function most likely drove its evolution. For example, the cytochrome $b_6f$ dimer interface is thought to form an electron transfer bridge for "cross talk" between the two monomers. In the case of rhodopsin, dimerization is thought to regulate G-protein coupling. Second, stability of newly evolved proteins must have been critical, and oligomerization may provide an efficient way to select for stabilizing mutants. In the case of a homodimer, for example, a single mutation at the interface could be twice as efficient in stabilizing the protein compared with a single mutation in a monomer. Similarly, a third explanation is that formation of oligomeric structures can augment genetic efficiency. For example, in the several ion channel structures we now know, identical subunits surround the ion pathway, requiring the coding of only a single unit to form a larger structure. As a fourth possible explanation, we speculate that if dense packing of membrane proteins is evolutionarily important for optimizing functional output per unit area of membrane, then the evolution of oligomeric interfaces might have been a structural adaptation to support high packing density while minimizing energetically unfavorable protein-protein contacts. Evolution would then have built upon this through adaptations that functionalized the interface.

Like oligomerization, the presence of strongly bound lipids and specific sites for water molecules found in recent high-resolution structures suggests a potentially large degree of structural, and hence functional, variability. What is their role in structural stability and what is their functional significance? In the case of $b_6f$, structural lipids are suggested either to be stabilizers, acting as "structural struts," or functional, "imposing restraints on protein dynamics." In the case of the ADP/ATP carrier, removal of structural cardiolipin molecules, as Nury et al. (2) point out, leads to a 20% decrease in protein activity. How dynamic is the lipid binding? Should these lipids be considered as ligands or as parts of the structure? In some cases it seems that the lipid requirement is not specific, but in others it is. Imagine the impact if it is found that these interactions with lipids are a highly regulated and ubiquitous cellular phenomenon. Similar questions can and should be asked for the fascinating case of structural waters, as well as for all posttranslational modifications, including glycosylation. The potential combinatorial explosion because of variable glycosylation patterns on the surface of membrane proteins underscores our contention, as we now discuss, that the size of structure space may be staggeringly large if all variations have functional relevance.

## WHY DYNAMICS?

In many cases, a protein must undergo a dynamic conformational transition between discrete structural states to carry out its function. Such transitions, for example from state A to state B, involve a change in the thermodynamic free energy of the system, $\Delta G_{A\to B}$. The ADP/ATP carrier protein located in the mitochondrial membrane, discussed by Nury et al. (2), is a good example. The authors point to the "induced transition fit" (ITF) mechanism to explain the dynamics of the protein. In this mechanism, described by Klingenberg (3), the membrane protein exists in multiple, discrete conformations, with the metabolite only binding perfectly to the highest energy state. The energy of this transition state is then utilized in triggering further conformational changes necessary for metabolite release. The total free energy change of such a process is the sum over all intermediate states:

$$\Delta G_{A\to B} = \sum_{i=1}^{N-1} \Delta G_{i\to i+1},$$

with $i = 1$ corresponding to state A, $i = N - 1$ corresponding to state B, and $N$ commonly assumed to be a finite and reasonably small number of states.

Just how many relevant energy states exist for any given mechanism? Five such states are suggested in the case of the ITF mechanism. In the case of ion channels, models built from gating behavior suggest an even larger number (4). If one thinks in terms of a multidimensional free energy landscape for transitions in proteins, then these discrete states are found either in local energy minima (intermediates) or maxima (transition states). However, the energy wells corresponding to the minima may be quite broad, and transitions between them may follow multiple paths. Each of these paths may, in fact, be populated with an extremely large number of functionally pertinent conformational substates. Such an idea is supported by the case of calmodulin, wherein analysis of disorder in the crystal structure has suggested that the protein "may sample a quasi-continuous spectrum of conformations" (5). Similarly, a very large number of paths between any two states are theoretically possible, although the accessible number is likely diminished by environmental conditions (see below). This all suggests that the overall free energy change may also be expressed as

$$\Delta G_{A\to B} = \sum_{j=1}^{N'-1} \Delta G'_{j\to j+1} = \sum_{k=1}^{N''-1} \Delta G''_{k\to k+1} = \cdots.$$

Which path is taken determines the work required in effecting a transition and thus has important consequences for protein efficiency and function. Could path choice be a variable parameter in cellular control, for example, through concerted variations in the proteins' surroundings (i.e., the membrane)?

Clearly, given the relevance of conformational states to function, single structures can only partially describe a mechanism. The range of available energy states along a transition path may be exploited by proteins for tuning function. How do we investigate the transition path? Is it possible to predict, solely from two static structures, the relevant path between the two states? The problem is especially complicated because structural dynamics span a range of frequencies, from low (large conformational changes or domain movements, more traditionally associated with function) to high (side-chain rotameric isomerizations, which may, for example, play an important role in electron transport proteins, which are otherwise relatively immobile). There is, however, significant progress being made in the area of computational biology that portends a numerical solution to this problem, although further improvements in the computational representation of chemistry are needed. For example, molecular dynamics simulations generally provide information about high-frequency fluctuations (6), and normal mode analysis can help predict the lower frequency paths (7). More sophisticated computational methods known as path sampling have been shown to yield highly detailed transition path information for simple systems (8–12), and more recently for biomolecules (13–16).

## ENVIRONMENT

As we suggested above, a high degree of complexity exists owing to structural variability and dynamic transition paths. A third factor that contributes to complexity is the membrane environment. Structure determination generally requires isolation of the protein from the remainder of the biological system. However, any proper thermodynamic analysis must include all relevant components of the system and must pay particularly close attention to boundaries where energy is exchanged. The contribution of the membrane environment therefore deserves consideration in analysis of function. We have noted the

multiplicity of available conformational transition paths, and that path choice determines the trade-off between energy spent and work done. The number of available paths, along with the paths themselves, must be altered by physical constraints placed on the protein by the membrane environment. Therefore, isolation of the protein from its native membrane provides only a partial story.

The membrane consists of, among other things, lipids and a multitude of proteins. In addition to severely biasing protein conformational states and the paths between them (17), the membrane itself is capable of storing energy through conformational flexibility of its own (18, 19). It has been clearly established that lipid composition varies profoundly in different membranes (20). Additionally, single membranes can be highly heterogeneous, as in the case of "lipid rafts," the functional manifestation of lipid domains known for decades to exist in synthetic mixtures (21–23). Sustaining such diversity, for example, through variable lipid composition, may cost the cell valuable resources, so specific functional rationales should be examined. Furthermore, biophysical measurements have established that physical properties of membranes, such as curvature elastic stress, profoundly affect the efficiency of membrane proteins. As one landmark example, a specific conformational transition in rhodopsin, the molecule discussed by Palczewki (24), is favored by the highly curved reverse hexagonal phase, rather than the standard lamellar phase (25). It seems likely that cells might take advantage of this type of specificity to tune function by genetically regulating their membrane-specific constituent lipid populations. Other significant membrane properties that are now appreciated as influencing structure and function of membrane proteins are, among others, lateral tension (26, 27), hydrophobic matching (28–32), and electrostatics (33, 34). By modifying these properties, cells have afforded themselves a highly tunable molecular environment and thus, we suggest, have gained incredibly fine control over protein function.

Molecular function underlies all of biology, from the processing of input in the simplest organisms to higher consciousness in humans. How intricate must the molecular machinery be to support such a diverse and elegant world? It is, of course, natural that we think in terms of discrete states. Primarily, it makes investigation of structure-function more tractable. Additionally, experimental techniques such as crystallization lend themselves to thinking in terms of single, or averaged, structures. The perspective we offer here poses a challenge to modern biologists given the immensity of combinatorial possibilities it suggests. It is hard to imagine how to begin to study this complexity; however, it is a challenge worth pursuing. If we seek a molecular-level explanation of phenomena as mystifying as consciousness, then our suggestion of near infinite combinatorial complexity seems less of a stretch. Clearly, the mere existence of complexity does not prove its evolutionary or functional significance. The important tasks are to digest the complexity and find the simplifications that remain true to the biology. We suggest that computational methods, along with increased experimental resolution, both spatially and temporally, should facilitate this effort. The progress reflected in the reviews that follow suggests that we are well on our way.



## ACKNOWLEDGMENTS

We thank the members of the Engelman lab and Daniel M. Zuckerman for helpful discussion. J.N.S. is supported by an N.R.S.A. Grant (GM071134-02) and D.M.E. by an N.I.H. grant (GM070895).

## LITERATURE CITED

1. Engelman DM. 2005. *Nature* 438:578–80
2. Nury AH, Dahout-Gonzalez C, Trézéguet V, Lauquin GJM, Brandolin G, Pebay-Peyroula E. 2006. *Annu. Rev. Biochem.* 75:713–41
3. Klingenberg M. 2005. *Biochemistry* 24:8563–70
4. Imredy JP, Yue DT. 1994. *Neuron* 12:1301–18
5. Wilson MA, Brunger AT. 2000. *J. Mol. Biol.* 301:1237–56
6. Crozier PS, Stevens MJ, Forrest LR, Woolf TB. 2005. *J. Mol. Biol.* 333:493–514
7. Valadie H, Lacapcre JJ, Sanejouand YH, Etchebest C. 2003. *J. Mol. Biol.* 332:657–74
8. Huber GA, Kim S. 1996. *Biophys. J.* 70:97–110
9. Pratt LR. 1986. *J. Chem. Phys.* 85:5045–48
10. Woolf TB. 1998. *Chem. Phys. Lett.* 289:433–41
11. Dellago C, Bolhuis PG, Csajka FS, Chandler D. 1998. *J. Chem. Phys.* 108:1964–77
12. Zuckerman DM, Woolf TB. 1999. *J. Chem. Phys.* 111:9475–84
13. Zuckerman DM. 2004. *J. Phys. Chem. B.* 108(16):5127–37
14. Bolhuis PG. 2005. *Biophys. J.* 88(1):50–61
15. Arora K, Schlick T. 2005 *J. Phys. Chem. B.* 109(11):5358–5367
16. Tobi D, Bahar I. 2005. *Proc. Nat. Acad. Sci.* 102(52):18908–13
17. Nanda H, Sachs JN, Petrache HI, Woolf TB. 2005. *J. Chem. Theory Comput.* 1:375–88
18. Mouritsen OG, Bloom M. 1993. *Annu. Rev. Biophys. Biomol. Struct.* 22:145–71
19. Hamill OP, Martinac B. 2001. *Physiol. Rev.* 81:685–740
20. Wenk MR, De Camilli P. 2004. *Proc. Natl. Acad. Sci. USA* 101:8262–69
21. Jain MK, White HB III. 1977. *Adv. Lipid Res.* 15:1–60
22. Klausner RD, Kleinfeld AM, Hoover RL, Karnovsky MJ. 1980. *J. Biol. Chem.* 255:1286–95
23. Simons K, van Meer G. 1988. *Biochemistry* 27:6197–202
24. Palczewki K. 2006. *Annu. Rev. Biochem.* 75:743–67
25. Brown MF. 1994. *Chem. Phys. Lipids* 73:159–80
26. Cantor RS. 1999. *Chem. Phys. Lipids* 101:45–56
27. Carrillo-Tripp M, Feller SE. 2005. *Biochemistry* 44:10164–69
28. Lee AG. 2003. *Biochim. Biophys. Acta* 1612:1–40
29. Williamson IM, Alvis SJ, East JM, Lee AG. 2002. *Biophys. J.* 83:2026–38
30. de Planque MRR, Killian JA. 2003. *Mol. Membr. Biol.* 20:271–84
31. Tieleman DP, Forrest LR, Sansom MSP, Berendsen HJC. 1998. *Biochemistry* 37:17554–61
32. Petrache HI, Zuckerman DM, Sachs JN, Killian JA, Koeppe RE, Woolf TB. 2002. *Langmuir* 18:1340–51
33. Aguilella VM, Bezrukov SM. 2001. *Eur. Biophys. J. Biophys. Lett.* 30:233–41
34. Jensen MO, Mouritsen OG. 2004. *Biochim. Biophys. Acta* 1666:205–26
35. Cramer WA, Zhang H, Yan J, Kurisu G, Smith JL. 2006. *Annu. Rev. Biochem.* 75:769–790

Annual Review
of Biochemistry
Volume 75, 2006

# Contents

# EXHIBIT A

# Dr Cameron Victor Jennings

1 Janice Street, Warners Bay NSW 2282
Australia
+61430095522
cam_jennings@hotmail.com

## EDUCATION:

**1. Post-Graduate Diploma in Energy Studies (Awarded in 2010)**
Murdoch University, School of Engineering and Energy
Perth, Australia

**2. Doctor of Philosophy (awarded in 2003)**
La Trobe University, Department of Biochemistry
Melbourne Australia.

**3. Bachelor of Science (First Class Honours– awarded in 1999)**
La Trobe University
Melbourne Australia.

## WORK EXPERIENCE:

**1. Senior Scientist/Project Leader (July 2007 – Present)**
**Immune System Therapeutics LTD. Sydney Australia.**
Role: Supervise and drive projects that increase the intellectual property profile of Immune System Therapeutics Ltd.

**2. Scientist (2006-July 2007)**
**Immune System Therapeutics LTD. Sydney Australia.**
Role: Drive projects that increase the intellectual property profile of the company

**3. Senior Post-Doctoral Research Fellow (2003-2006)**
**Harvard University, School of Public Health,**
**Department of Immunology and Infectious Diseases, Boston USA.**
Role: Perform and supervise research into Malaria vaccine candidates

**4. Laboratory class teaching assistant - Undergraduate Biochemistry (1999-2001)**
**La Trobe University, Department of Biochemistry. Melbourne, Australia.**
Role: Supervise undergraduate students in practical classes

**5. Research Assistant, Phillips Laboratory (1997-1999).**
**La Trobe University, Department of Biochemistry. Melbourne, Australia.**
Role: Support research into potential anti-cancer compounds

## OTHER ACADEMIC TRAINING:

**1. Environmental Auditor Certification Workshop (July, 2009)**
**Thomson - Reuters**
Aim: Training and certification for auditing environmental management systems

**2. Environmental Management Systems Workshop (July, 2009)**
**Thomson - Reuters**
Aim: Additional training in environmental management systems and auditing

**3. Integrated Management Systems Workshop (July, 2009)**
**Thomson - Reuters**
Aim: Training in integrating environmental management systems into existing business structures

**4. NSW Enterprise Workshop (Winter Program 2008)**
Aim: To develop and defend a business plan, including an executive summary, marketing, operations, project management and a financial plan

## ADDITIONAL ACHIEVEMENTS:

- Won a **business plan award** at the NSW Enterprise Workshop (Best Intellectual Property Plan, shared award with 2 other team members) 2008
- Supervised an honours student who received the **Deans Merit Award** at the University of Technology, Sydney 2008
- Jennings et al., 2005. was among the 20 most cited articles (position 11) of 2005, January – June.
- Awarded a monetary prize for the **best oral seminar** at the inaugural Melbourne Protein Group Symposium. 2002.
- Received a **student bursary** to attend the Forth Australian Peptide Conference (Lindeman Island, Queensland, Australia). 2001.
- **President of the La Trobe University Biochemistry Society**. 2000
- Accepted into the **Golden Key Club** for Scholastic achievements. 1997.
- Accepted onto the **Deans List** for Scholastic achievements (La Trobe University, Melbourne, Australia. 1997.

## PATENTS:

Novel nucleic acid molecules
International application WO 0134829 filed 25th November 2000.
**Inventors:** Craik, D.J., Anderson, M.A. and **Jennings, C.V.**

**Several Provisional patent applications written and filed**

## INVITED SEMINARS/ORAL PRESENTATIONS:

- **Institute for the Biotechnology or Infectious Diseases (University of Technology Sydney) 2007**
  Jennings C.V., Ahouidi A.D., Bei A., Sarr O., Ndir O., Wirth D., Mboup S., Duraisingh M.T

- **Biological Sciences and Public Health Retreat (Harvard University) 2005**
  Jennings C.V., Ahouidi A.D., Bei A., Sarr O., Ndir O., Wirth D., Mboup S., Duraisingh M.T.

- **Biological Sciences and Public Health Retreat (Harvard University) 2003**
  Jennings, C.V., Whitehurst, N., Desimone, A., Duraisingh, M.T.

- **PhD Qualification Seminar 2002**
  Jennings C.V. Craik D.J., Anderson M.A

- **Inaugural Melbourne Protein Group Symposium 2002**
  Jennings C.V., West J.A., Craik D.J., Anderson M.A.

- **PhD Conversion Seminar 2001**
  Jennings C.V. Craik D.J., Anderson M.A

## RESEARCH PUBLICATIONS:

Bushkin GG, Ratner DM, Cui J, Banerjee S, Duraisingh MT, **Jennings CV**, Dvorin JD, Gubbels MJ, Robertson SD, Steffen M, O'Keefe BR, Robbins PW, Samuelson J.
Eukaryot Cell. 2009 Oct 16.

Desimone TM*, **Jennings CV***, Bei AK, Comeaux C, Coleman BI, Refour P, Triglia T, Stubbs J, Cowman AF, Duraisingh MT.
Mol Microbiol. 2009 EPub. Apr 14
* Joint first authors

Lantos PM, Ahouidi AD, Bei AK, **Jennings CV**, Sarr O, Ndir O, Wirth DF, Mboup S, Duraisingh MT.
Parasitology. 2009 Jan;136(1):1-9.

Desimone TM, Bei AK, **Jennings CV**, Duraisingh MT.
Int J Parasitol. 2008 Mar;39(4):399-405.

Duraisingh MT, DeSimone T, **Jennings C**, Refour P, Wu C
Subcell Biochem 2008.:46-57.

Gillon AD, Saska I, **Jennings CV**, Guarino RF, Craik DJ, Anderson MA.
Plant J. 2008 Feb;53(3):505-15.

**Jennings CV**, Ahouidi AD, Zilversmit M, Bei A, Rayner J, Sarr O, Ndir O, Wirth DF, Mboup S, Duraisingh MT.
Infection and Immunity. 2007 Jul;75(7):3531-3538.

**Jennings C.V.**, Rosengren K.J., Daly N.L., Plan M., Stevens J., Scanlon M.J., Waine C., Norman D.G., Anderson M.A., Craik D.J.
Biochemistry. 2005 Jan 25;44(3):851-860.

Dutton J.L., Renda R.F., Waine C., Clark R.J., Daly N.L., **Jennings C.V.**, Anderson M.A., Craik D.J.
J Biol Chem. 2004 Nov 5;279(45):46858-46867.

**Jennings C.V.**, West J., Waine C., Craik D., Anderson M.
Proc. Natl. Acad. Sci. USA. 2001 Sep 11;98(19):10614-10619.

Craik, D.J., Anderson, M.A., Barry, D.G., Clark, R.J., Daly, N.L., **Jennings, C. V.**, Mulvenna, J.
International Journal of Peptide Research and Therapeutics, 2001 May;8(3-5): 119-128.